Supplement to 'Prognostic value of genomic signatures in metastatic Clear Cell Renal Cell Carcinoma using The Cancer Genome Atlas data'

de Velasco G, Culhane AC et al.,

May 23, 2016

Contents

Reproducibility, code, data

All statistical analyses were performed in the R statistical language. All code and data to reproduce these analyses are available (CC-BY).

Obtaining, processing TCGA ccRCC RNAseq Data

TCGA RNA sequencing data were downloaded from https://tcga-data.nci.nih.gov/tcgafiles/ftp_auth/distro_ftpusers/anonymous/tumor/kirc/cgcc/unc.edu/illuminahiseq_rnaseqv2/rnaseqv2/unc.edu_KIRC.IlluminaHiSeq_RNASeqV2.Level_3.1.6.0/

Files (n=606) that ended in ".rsem.genes.normalized_results" were used.

These are normalized to a fixed upper quartile value of 1000 for gene level estimates. Only primary tumors with tissue code TP (n=533) were retained. RNAseq of tissue from a second primary (Code TAP; n=1) or normal (Code NT; n=72) tissue were excluded.

Tissue Codes

```
ID  Description                Code
1   Primary solid Tumor        TP
5   Additional - New Primary   TAP
```

```
11  Solid Tissue Normal         NT
```

Counts of Tumors with ID and Tissue Codes

```
##
##       01A 01B 05A 11A
##   NT    0   0   0  72
##   TAP   0   0   1   0
##   TP  529   4   0   0
```

Data were transformed to log2+1 values. In analysis below, rows were also median centered.

R Expression Set Information

```
## ExpressionSet (storageMode: lockedEnvironment)
## assayData: 20531 features, 533 samples
##   element names: exprs
## protocolData: none
## phenoData
##   sampleNames: TCGA-BP-4342-01A-01R-1289-07
##     TCGA-CZ-4862-01A-01R-1305-07 ... TCGA-CJ-4881-01A-01R-1305-07
##     (533 total)
##   varLabels: barcode SampleID ... sampleType (10 total)
##   varMetadata: labelDescription
## featureData
##   featureNames: ?|100130426 ?|100133144 ... tAKR|389932 (20531
##     total)
##   fvarLabels: SYMBOL ENTREZID
##   fvarMetadata: labelDescription
## experimentData: use 'experimentData(object)'
## Annotation:
```

Clinical information on patients with metastatic RCC in the TCGA

Among the 533 patients with RNAseq profiles of their primary tumors, in review of clinical records we identified 56 patients who had metastatic disease in the TCGA ccRCC cohort.

mRCC IMDC and MSKCC subtypes

Among the 56 patients, their IMDC and MSKCC prognostic risk groups were favorable (8/8), intermediate (37/40), poor (11/8) respectively.

The same 8 patients had favorable risk in both the IMDC and MSKCC risk scores, however there were differences in risk classification to intermediate and poor risk groups. A cross table of the IMDC and MSKCC risk score classifications was;

```
##               IMDC
## MSKCC          favorable intermediate poor
##   favorable            8            0    0
##   intermediate         0           34    6
##   poor                 0            3    5
```

The median survival (in months) for each risk group is below:

```
## [1] "MSKCC"
```

```
##    MSKCC=favorable MSKCC=intermediate         MSKCC=poor
##           98.00548           22.29041           23.12877
```

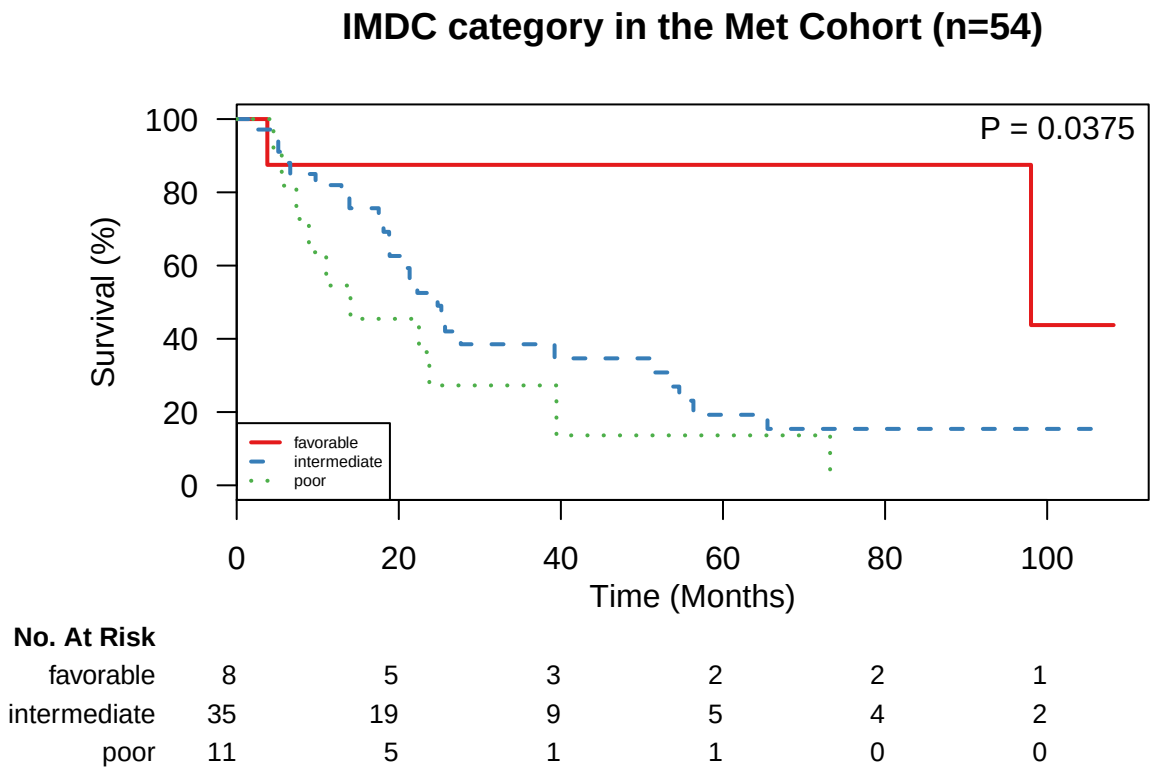
```
## [1] "IMDC"
```

```
##    IMDC=favorable IMDC=intermediate         IMDC=poor
##          98.00548          24.78904          14.03836
```

Both the IMDC and MSKCC risk criteria were significantly associated with outcome ($p<0.05$). There was a significantly greater risk of death in patients assigned to the poor outcome group compared to the favorable risk group in both the IMDC and MSKCC classifications. The intermediate subtypes of either MSKCC or IMDC were not significantly different to the favorable group.

Assocation of IMDC risk criteria and surival in mRCC patients

```
## Call:
## coxph(formula = Surv(TIME, EVENT) ~ IMDC, data = pData(TCGAMet))
##
##   n= 54, number of events= 37
##    (2 observations deleted due to missingness)
##
##                     coef exp(coef) se(coef)     z Pr(>|z|)
## IMDCintermediate 1.3278    3.7728   0.7398 1.795   0.0727 .
## IMDCpoor         1.8290    6.2275   0.7851 2.329   0.0198 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                  exp(coef) exp(-coef) lower .95 upper .95
## IMDCintermediate     3.773     0.2651    0.8851     16.08
## IMDCpoor             6.227     0.1606    1.3366     29.02
##
## Concordance= 0.601  (se = 0.046 )
## Rsquare= 0.13   (max possible= 0.988 )
## Likelihood ratio test= 7.54  on 2 df,   p=0.02303
## Wald test            = 5.71  on 2 df,   p=0.05767
## Score (logrank) test = 6.56  on 2 df,   p=0.03754
```

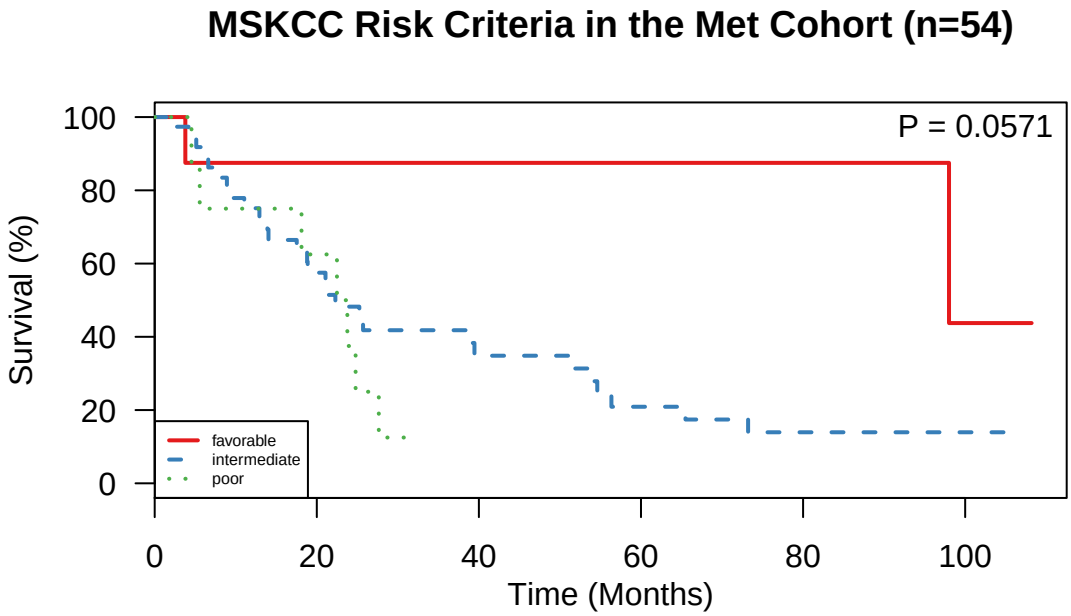


No. At Risk						
favorable	8	5	3	2	2	1
intermediate	35	19	9	5	4	2
poor	11	5	1	1	0	0

Figure 1: Survival analysis of IMDC risk criteria in mRCC

Assocation of MSKCC risk criteria and surival in mRCC patients

```
## Call:
## coxph(formula = Surv(TIME, EVENT) ~ MSKCC, data = pData(TCGAMet))
##
##   n= 54, number of events= 37
##    (2 observations deleted due to missingness)
##
##                      coef exp(coef) se(coef)     z Pr(>|z|)
## MSKCCintermediate 1.3739    3.9508   0.7365 1.865   0.0621 .
## MSKCCpoor         1.8291    6.2285   0.8212 2.227   0.0259 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                   exp(coef) exp(-coef) lower .95 upper .95
## MSKCCintermediate     3.951     0.2531    0.9327     16.73
## MSKCCpoor             6.228     0.1606    1.2456     31.14
##
## Concordance= 0.575  (se = 0.044 )
## Rsquare= 0.12   (max possible= 0.988 )
## Likelihood ratio test= 6.9  on 2 df,   p=0.03174
## Wald test            = 4.96  on 2 df,   p=0.08363
## Score (logrank) test = 5.73  on 2 df,   p=0.05709
```



No. At Risk						
favorable	8	5	3	2	2	1
intermediate	38	19	10	6	4	2
poor	8	5	0	0	0	0

Figure 2: Survival analysis of MSKCC risk criteria in mRCC

Annotating genes in the ClearCode34 Signature

The ClearCode34 signature (Brooks et al., 2014) is a list of 34 genes, that separates ccRCC into two groups, ccA and ccB. It was developed from several gene lists;

- a list of 120 genes (Brandon et al., 2010),
- gene signatures that were shown to be differentially expressed between the ccA/ccB subtypes using significance analysis of microarrays
- and other published markers (Ellis et al., 2006, Goel et al., 2011, Harris et al., 2002, Wright et al., 2009, Yao et al., 2008).

In the Brooks et al., 2014 paper, these gene lists, the ClearCode34 and other gene lists from which it was developed are supplied in Table 3 and Supplemental Table 1 respectively. In order to annotate (as far as possible) the ClearCode34 genes, These were downloaded and extracted from the following tables from publications Brooks et al., and Brandon et al.,

- Table 3 of Brooks et al., The ClearCode34 genes (`24613583_Table3.txt`)
- A Table containing the probeID and source of the 120 genes provided in Supplementary Table 5 from Brandon et al., 2010 (`20871783_Brannon_2010_SupplTable5_120probes.txt`).

GeneList Source of the ClearCode34 genes

The 120 probe Brandon list contained the source AgilentID for each gene and therefore 23/34 genes are in the 120 probelist table. There is no Agilent (source identifier) for the remaining genes; 4 genes from the SAM list and 7 from the lists of prognostic markers.

Using the list of AgilentID, gene symbols were updated using the Bioconductor annotation package hgug4112a.db. Two genes had updated gene symbols, these were C13orf1 and UNG2. These are now SPRYD7 and CCNO respectively

The updated annotation for the ClearCode34 signature is provided in Table 1.

Table 1: ClearCode34 Gene Signature

	Subtype	Accession	Brandon_120	SAM_8	Prognosis_12	SYMBOL
MAPT	ccA	NM_016835	TRUE	FALSE	FALSE	MAPT
STK32B	ccA	NM_018401	TRUE	FALSE	FALSE	STK32B
FZD1	ccA	NM_003505	TRUE	FALSE	FALSE	FZD1
RGS5	ccA	NA	FALSE	FALSE	TRUE	RGS5
GIPC2	ccA	NM_017655	TRUE	FALSE	FALSE	GIPC2
PDGFD	ccA	NM_025208	TRUE	FALSE	FALSE	PDGFD
EPAS1	ccA	NA	FALSE	FALSE	TRUE	EPAS1
MAOB	ccA	NM_000898	TRUE	FALSE	FALSE	MAOB
CDH5	ccA	NA	FALSE	FALSE	TRUE	CDH5
TCEA3	ccA	NM_003196	TRUE	FALSE	FALSE	TCEA3
LEPROTL1	ccA	NM_015344	TRUE	FALSE	FALSE	LEPROTL1
BNIP3L	ccA	NM_004331	TRUE	FALSE	FALSE	BNIP3L
EHBP1	ccA	NM_015252	TRUE	FALSE	FALSE	EHBP1
VCAM1	ccA	NA	FALSE	FALSE	TRUE	VCAM1
PHYH	ccA	NM_006214	TRUE	FALSE	FALSE	PHYH
PRKAA2	ccA	NM_006252	TRUE	FALSE	FALSE	PRKAA2
SLC4A4	ccA	NM_003759	TRUE	FALSE	FALSE	SLC4A4
ESD	ccA	NM_001984	TRUE	FALSE	FALSE	ESD
TLR3	ccA	NM_003265	TRUE	FALSE	FALSE	TLR3
NRP1	ccA	NA	FALSE	FALSE	TRUE	NRP1
C11orf1	ccA	NM_022761	TRUE	FALSE	FALSE	C11ORF1
ST13	ccA	NM_003932	TRUE	FALSE	FALSE	ST13
ARNT	ccA	NA	FALSE	FALSE	TRUE	ARNT
C13orf1	ccA	NM_020456	TRUE	FALSE	FALSE	SPRYD7
SERPINA3	ccB	NA	FALSE	TRUE	FALSE	SERPINA3
SLC4A3	ccB	NA	FALSE	TRUE	FALSE	SLC4A3
MOXD1	ccB	NA	FALSE	TRUE	FALSE	MOXD1
KCNN4	ccB	NM_002250	TRUE	FALSE	FALSE	KCNN4
ROR2	ccB	NA	FALSE	FALSE	TRUE	ROR2
FLJ23867	ccB	AK074447	TRUE	FALSE	FALSE	FLJ23867
FOXM1	ccB	NA	FALSE	TRUE	FALSE	FOXM1
UNG2	ccB	NM_021147	TRUE	FALSE	FALSE	CCNO
GALNT10	ccB	AK021777	TRUE	FALSE	FALSE	GALNT10
GALNT4	ccB	NM_003774	TRUE	FALSE	FALSE	GALNT4

The ClearCode34 Model: Building and reproducing published results

We ensure we correcly implemented the ClearCode34 Model, we built and reproduced two results described in Brooks et al., (1). We reproduced classification of 153 UNC patients and 380 TCGA samples described in Supplementary Tables 4 and 3 published by Brooks et al., (2012) (PMID: 24613583)

Training and implementing the ClearCode34 model

The ClearCode34 model described by Brooks et al., (1), distinguishes two subtypes of RCC (ccA and ccB), and was trained using a nearest centroid classifier (using the R package pamr). Author of Brooks et al., 2014 kindly sent us training data and code (in the R statistical language) to build the ClearCode34 pamr classifier and replicate their analysis. The training data contained gene expression profiles of the ClearCode 34 genes in 40 samples, of which 23 were ccA and 17 were ccB cases.

```
## [1] "Building ClearCode34 Model"
```

```
## 12345678910111213141516171819202122232425262728293O
```

Succesfully reproducing classification of 153 UNC patients described in Supplementary Table 4 by Brooks et al., (2012)

The authors also provided us the UNC test data cohort of the 34 gene expression profiles in 167 tumors of UNC patients. The ccA/ccB classification of 153 of these patients were available in Supplementary Table 4 Brooks et al., (2012). Therefore using the ClearCode34 classifier model, we classified ccA/ccB subtype of the 153 tumors.

All cases were correctly assigned (ccA n=67; ccB n=86) demonstrating we have correctly implemented the ClearCode34 classifier algorithm.

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction ccA ccB
##        ccA  67   0
##        ccB   0  86
##
##                Accuracy : 1
##                  95% CI : (0.9762, 1)
##     No Information Rate : 0.5621
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 1
##  Mcnemar's Test P-Value : NA
##
##             Sensitivity : 1.0000
##             Specificity : 1.0000
##          Pos Pred Value : 1.0000
##          Neg Pred Value : 1.0000
##              Prevalence : 0.4379
##          Detection Rate : 0.4379
##    Detection Prevalence : 0.4379
##       Balanced Accuracy : 1.0000
```

```
##
##        'Positive' Class : ccA
##
```

The ccA, ccB probability results were identical to Supplementary Table 4 from Brooks et al. 2012. (Pearson Correlation Coefficient of 1, p value =0)

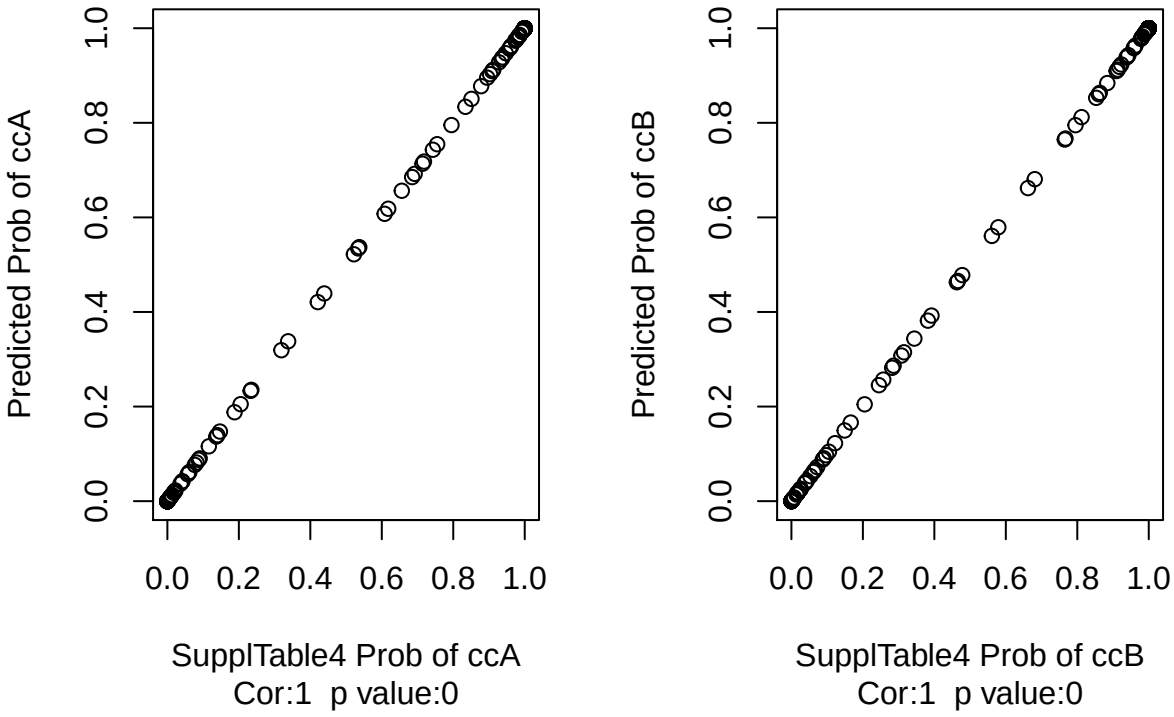


Figure 3:

Reproducing classification of 380 TCGA samples described in Supplementary Tables 3 published by Brooks et al., (2012)

To further check our implementation of the ccA/ccB classifier, we classified 380 TCGA samples and compared these to classifications that Brooks et al.2014 published in European Urology (in supplementary file 3 (24613583_Table3.txt).

During this analysis, we observed that the classification of the 380 cohort was sensitive to scaling and median centering of the data. Therefore we dsecribe a detailed exploration of this first.

Effect of median scaling of genes on performance of ClearCode34 classifier

As part of the ClearCode algorithm, data must be logged (log2), median centered and scaled. However the median gene value maybe poorly estimated if the sample size is low and is dependent on the relative proportion of ccA/ccB cases.

The effects of scaling on these data are considerale, for example without row centering, the accuracy was only 67%. Median-centering genes (rows) greatly improved the specificity.

We describe the effect of scaling/centering on classification of 380 TCGA tumors classified as ccA (n=205) and ccB (n=175) by Brooks et al., manuscript (Supplementary Table 3)

1. **No scaling/row centering**. When no scaling or median centering is performed all tumors are assigned to ccA. Therefore 175 ccB samples are incorrectly classified.

```
predict(ClearCode34.model,newdata)
```

```
##          Brooks et al.,
## predicted ccA ccB
##       ccA 205 175
```

2. **Scaling only**. (column centering/scaling, that is subtract mean, divide by the standard deviation); Most (n=321) tumors are classified as ccA. 116 patients are incorrectly classified as ccA when they should be ccB.

```
scaledData<-scale(newdata)
predict(ClearCode34.model,scaledData)
```

```
##          Brooks et al.,
## predicted ccA ccB
##       ccA 205 116
##       ccB   0  59
```

3. **Scale to the median of the 533 cohort (complete primary tumor TCGA dataset)**. With increased sample size, there is a greater chance of estimating the true median gene expression value. Whilst the median of the larger data set would not have been available in the Brooks et al., study of 380. Using the median of the 533 cohort, we classify 201 ccA and 179 ccB cases. There were 4 mis-classifieds samples; we misclassify 4 ccA samples as ccB

```
medianScaledData<-sweep(newdata, 1, rowMedians(TCGA533))
scaledData<-scale(medianScaledData)
predict(ClearCode34.model,scaledData)
```

```
##          Brooks et al.,
## predicted ccA ccB
##       ccA 201   0
##       ccB   4 175
```

4. **Scale to the median of the 380 cohort** Using the median of the 380 cohort, we classify 191 ccA and 189 ccB cases. There are 14 misclassified ccA samples. 14 ccA samples as predicted to be ccB

```
medianScaledData<-sweep(newdata, 1, rowMedians(TCGA380))
scaledData<-scale(medianScaledData)
predict(ClearCode34.model,scaledData)
```

```
##          Brooks et al.,
## predicted ccA ccB
##       ccA 191   0
##       ccB  14 175
```

5. **Scale to the median of the Metastatic (n=56) cohort** Using the median of the metastatic cohort. When we use the gene median of the mRCC cases, we classify 263 ccA and 117 ccB. There were 58 misclassified samples. 58 ccB samples as ccA.

```
medianScaledData<-sweep(newdata, 1, rowMedians(TCGA56))
scaledData<-scale(medianScaledData)
predict(ClearCode34.model,scaledData)
```

```
##          Brooks et al.,
## predicted ccA ccB
##       ccA 205  58
##       ccB   0 117
```

These data are summarized in Table 2.

Table 2: Effect of median scaling on Clearcode34 performance

	none	scale_only	scale_Median533	scale_Median380	scale_MedianMet56
Sensitivity	1.000	1.000	0.980	0.932	1.000
Specificity	0.000	0.337	1.000	1.000	0.669
Pos Pred Value	0.539	0.639	1.000	1.000	0.779
Neg Pred Value	NaN	1.000	0.978	0.926	1.000
Prevalence	0.539	0.539	0.539	0.539	0.539
Detection Rate	0.539	0.539	0.529	0.503	0.539
Detection Prevalence	1.000	0.845	0.529	0.503	0.692
Balanced Accuracy	0.500	0.669	0.990	0.966	0.834

Therefore to obtain robust classifications, we used constant scaling factor to median center genes. This was the median of the 533 cohort.

Reproducing classification of 380 TCGA samples described in Brooks et al Supplementary Tables 3 with 99% accuracy

The TCGA data was log 2 transformed (log2 +1) and the 34 genes were median centered about the median of 533 tumors expression values. We predicted all 533 TCGA cases, of these 268 were ccA and 265 were ccB. Compared these prediction to the 380 TCGA tumors predicted by Brooks et al.,

In the Brooks et al., 380 TCGA tumors were classified as ccA (n=205) and ccB (n=175). Using our implementation of the ClearCode34 alogorithm we predicted these 201 and 179 tumors to be ccA and ccB respectively. There were 4/380 mis-classified cases, giving a classifier accuracy of 99%, with 98% Sensitivity and 100% Specificity. This compared well to results presented by Brooks et al., in Suppl Table 3.

```
## Confusion Matrix and Statistics
##
```

```
##           Reference
## Prediction ccA ccB
##        ccA 201   0
##        ccB   4 175
##
##                Accuracy : 0.9895
##                  95% CI : (0.9733, 0.9971)
##     No Information Rate : 0.5395
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.9789
##  Mcnemar's Test P-Value : 0.1336
##
##             Sensitivity : 0.9805
##             Specificity : 1.0000
##          Pos Pred Value : 1.0000
##          Neg Pred Value : 0.9777
##              Prevalence : 0.5395
##          Detection Rate : 0.5289
##    Detection Prevalence : 0.5289
##       Balanced Accuracy : 0.9902
##
##        'Positive' Class : ccA
##
```

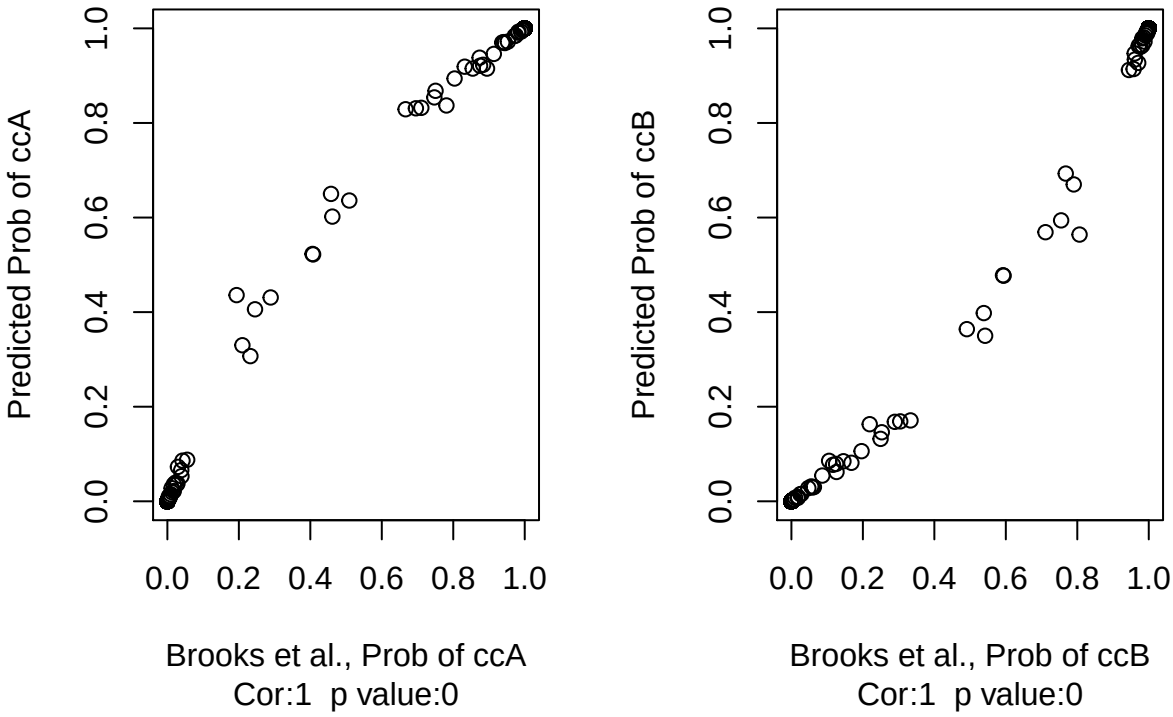


Figure 4:

ClearCode34 ccA/ccB Subtype classification of Metastatic RCC cases (n=56)

We identified 57 metastatic patients in the TCGA cohort, but we only had RNA sequencing data for 56 cases. These came from 5 institutions.

```
##
##     B0 B8 BP CJ CZ
##   1  0  0  0  0 11
##   2  0  0  0 20  0
##   3  0  0 11  0  0
##   4 12  0  0  0  0
##   5  0  2  0  0  0
```

Applying the ClearCode34 classifier, the mRCC patient classified as: 18 ccA, 38 ccB.

The larger group of ccB tumors (n=38) had worse overall survival (median survival time of 22.3 vs 27.65 months):

```
##
## ccA ccB
##  18  38
```

```
## prediction=ccA prediction=ccB
##       27.64932       22.29041
```

Among these 56 patients, a subset of these cases (n=22) were among the 380 tumors predicted by Brooks et al., The predictions of these 22 cases were identical between Brooks et al., and our prediction correctly (8 ccA and 14 ccB cases, 100% concordant)

Clearcode34 ccA/ccB subtypes and IMDC (or MSKCC) risk criteria class are distinct

In mRCC the IMDC or MSKCC risk critiera are widely used to predict patient outcome. Therefore we explored whether the ccA/ccB provided new information.

Whilst the IMDC and MSKCC classifications have highly significant overlap (p value <0.00001). Both risk criteria classify the same 8 patients as favorable. They differ in which casses are classified as intermediate or poor.

```
##
##  Pearson's Chi-squared test
##
## data:  TCGAMet$MSKCC and TCGAMet$IMDC
## X-squared = 65.934, df = 4, p-value = 1.635e-13
```

We observed no significant overlap in the ccA/ccB subtype classification and the IMDC or IDMC risk class (p =0.8633, p=0.863 respectively), indicating that the ccA/ccB classification provides new and different information to existing risk criteria in mRCC

Whilst the IMDC and MSKCC identify the same 8 patients as favorable. Only 3 /8 cases have favorable ClearCode34 classification as ccA. 5/8 favorable. cases are predicted to have poor outcome (ccB). Similarly 4/11 IMDC and 3/8 MSKCC poor risk criteria groups are predicted to have good outcome (ccA) status using the ClearCode34 classifier.

```
## [1] "Overlap with IMDC"
```

```
##
##       favorable intermediate poor
##   ccA         3           11    4
##   ccB         5           26    7
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(TCGAMet$prediction, TCGAMet$IMDC)
## X-squared = 0.29389, df = 2, p-value = 0.8633
```

```
## [1] "Overlap with MSKCC"
```

```
##
##       favorable intermediate poor
##   ccA         3           12    3
##   ccB         5           28    5
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(TCGAMet$prediction, TCGAMet$MSKCC)
## X-squared = 0.29474, df = 2, p-value = 0.863
```

Gene Expression of the 34 genes in mRCC (Heatmap)

Data were log2 transformed (log2+1) and median centered (using the row medians of the 533 data set) as above. Unsupervised hierarchical cluster analysis (distance was 1- Pearson Correlation Coefficient, with average linkage) was applied to the gene expression (RNAseq) profiles of the 34 genes in the 56 mRCC tumor and to the entire n=533 tumors.

Note tumors the ccA/ccB classification and MSKCC and IMDC classification are distinct.

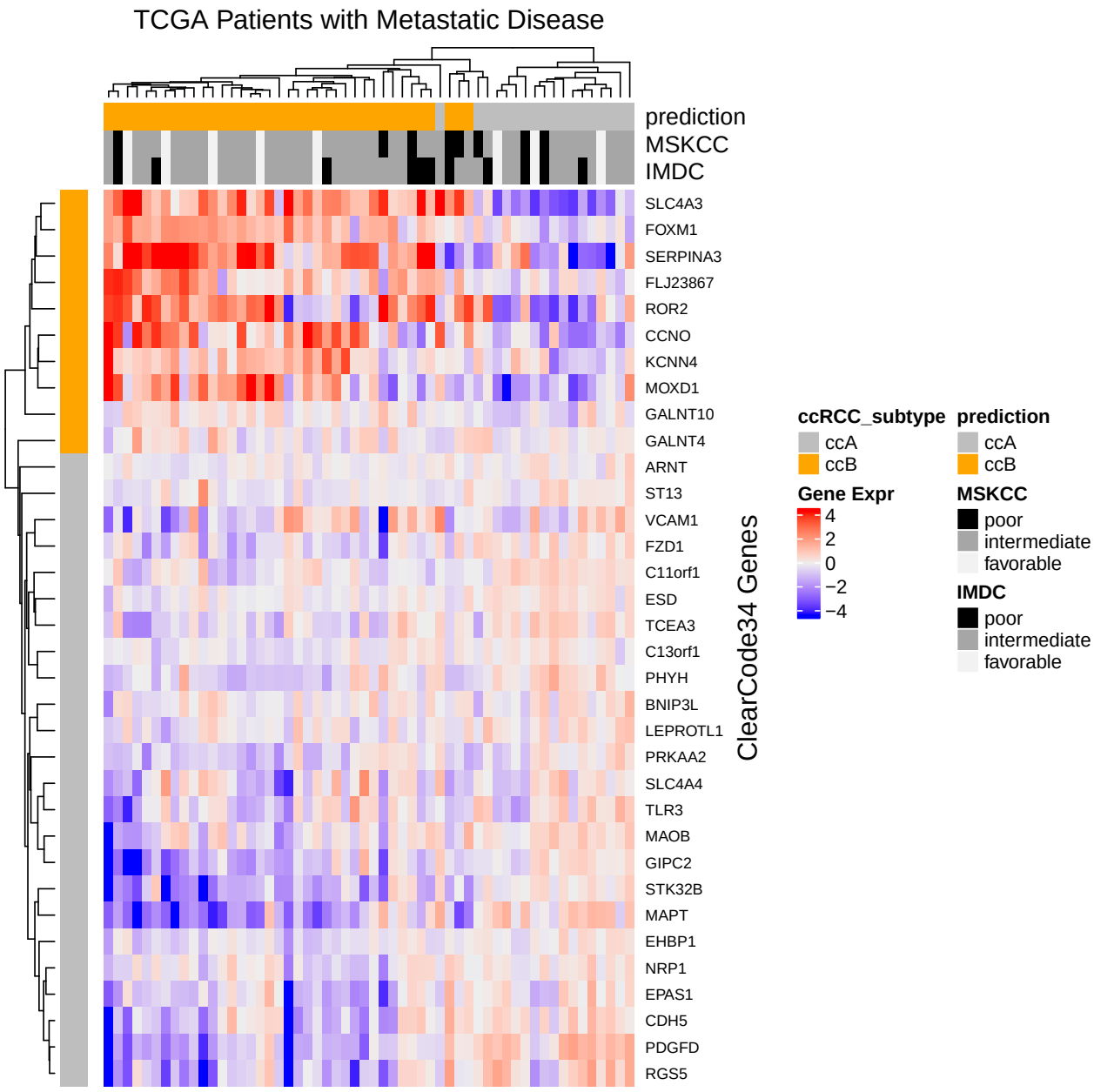


Figure 5: Gene Expression of ClearCode34 genes in mRCC

The ClearCode34 genes did not distinguish mRCC from non-metastatic RCC

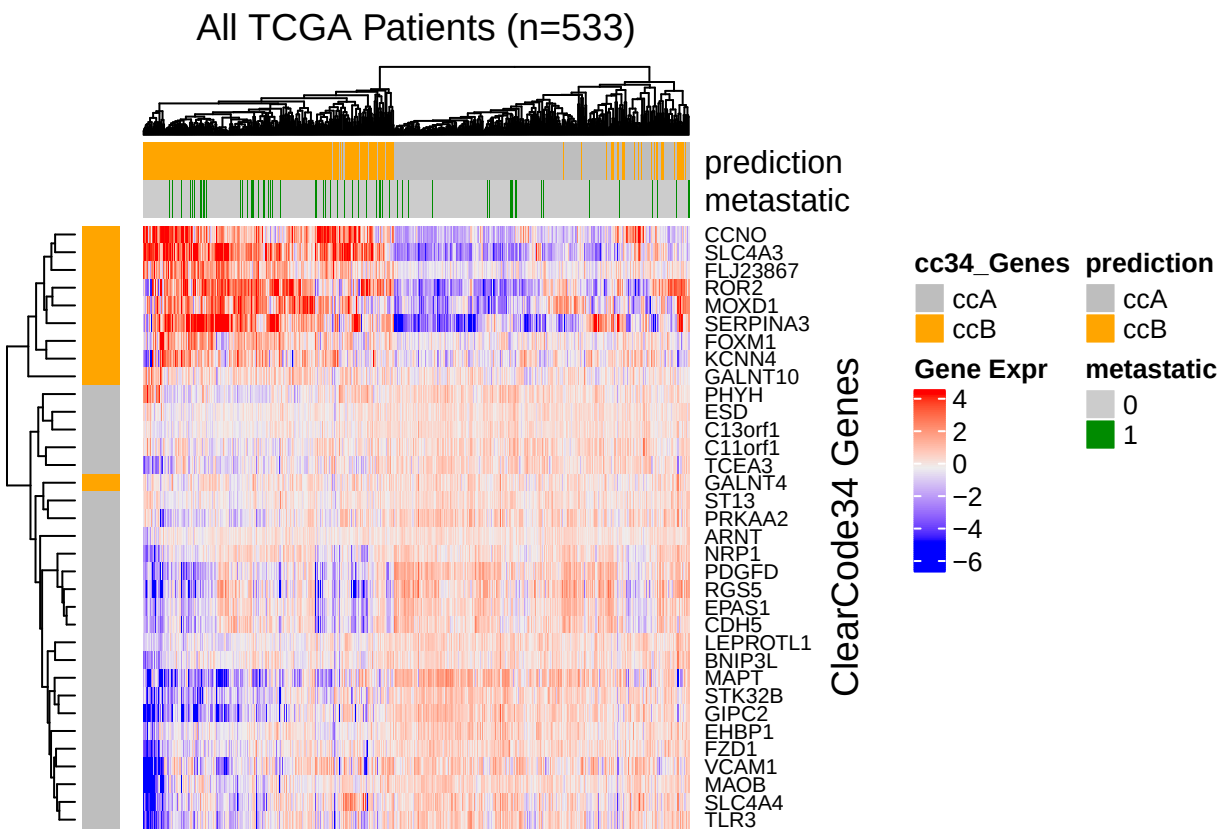


Figure 6: Gene Expression of ClearCode34 genes in all TCGA RCC cases (primary and met)

ClearCode34 ccB subtype has poorer overall survival in mRCC

There was gene expression (RNAseq) data for 56 mRCC, but we had survival data on n=54 patients. These 54 mRCC patient had ccA (n=17) /ccB (n=37) ClearCode34 classification.

Patients with ccA subtype (n=17) had a median overall survival of 27.6 months, compared to patients with ccB subtype (n=37) who had a median overall survival of 22.3 months.

```
## Call: survfit(formula = Surv(TIME, EVENT == 1) ~ prediction, data = pData(TCGAMet))
##
##    2 observations deleted due to missingness
##                 n events median 0.95LCL 0.95UCL
## prediction=ccA 17      7   27.6    22.5      NA
## prediction=ccB 37     30   22.3    14.0    51.5
```

Results of Cox PH survival analysis of mRCC (n=54) show that ccB patients had significantly worse overall survival compared to ccA mRCC patients (HR 2.33, CI 1.02-5.31, p <0.05).

The Hazard Ratio for the ClearCode34 model ccA/ccB subtypes was exp(0.844) which is a hazard Ratio 2.32, indicating that ccB have a higher risk of death than ccA (p-value for ccB is p=0.045). The p-values for three alternative tests for overall significance of the model is given by the Likelihood ratio test (p=0.030), Wald test (p=0.045) and log rank Test (p=0.039) on 1 df.

```
## Call:
## coxph(formula = Surv(TIME, EVENT == 1) ~ prediction, data = pData(TCGAMet))
##
##   n= 54, number of events= 37
##    (2 observations deleted due to missingness)
##
##                 coef exp(coef) se(coef)     z Pr(>|z|)
## predictionccB 0.8444    2.3265   0.4211 2.005    0.045 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
## predictionccB     2.326     0.4298     1.019     5.311
##
## Concordance= 0.583  (se = 0.045 )
## Rsquare= 0.083   (max possible= 0.988 )
## Likelihood ratio test= 4.71  on 1 df,   p=0.03005
## Wald test            = 4.02  on 1 df,   p=0.04497
## Score (logrank) test = 4.26  on 1 df,   p=0.03903
```

```
##                                predictionccB
## hazard.ratio                      2.32648346
## coef                              0.84435788
## se_coef                           0.42114331
## lower95                           1.01911251
## upper95                           5.31101840
## coef.pvalue                       0.04497179
## Score_logrank.p.value.pvalue      0.03902867
## Wald.p.value.pvalue               0.04497179
## LikelihoodRatio.p.value.pvalue    0.03005286
```

The Clearcode34 ccA/ccB, MSKCC and IMDC risk criteria all predict overall survival in mRCC. There we compared the prognostic performance (univariate models) of each using Cox PH and concordance index survival analysis.

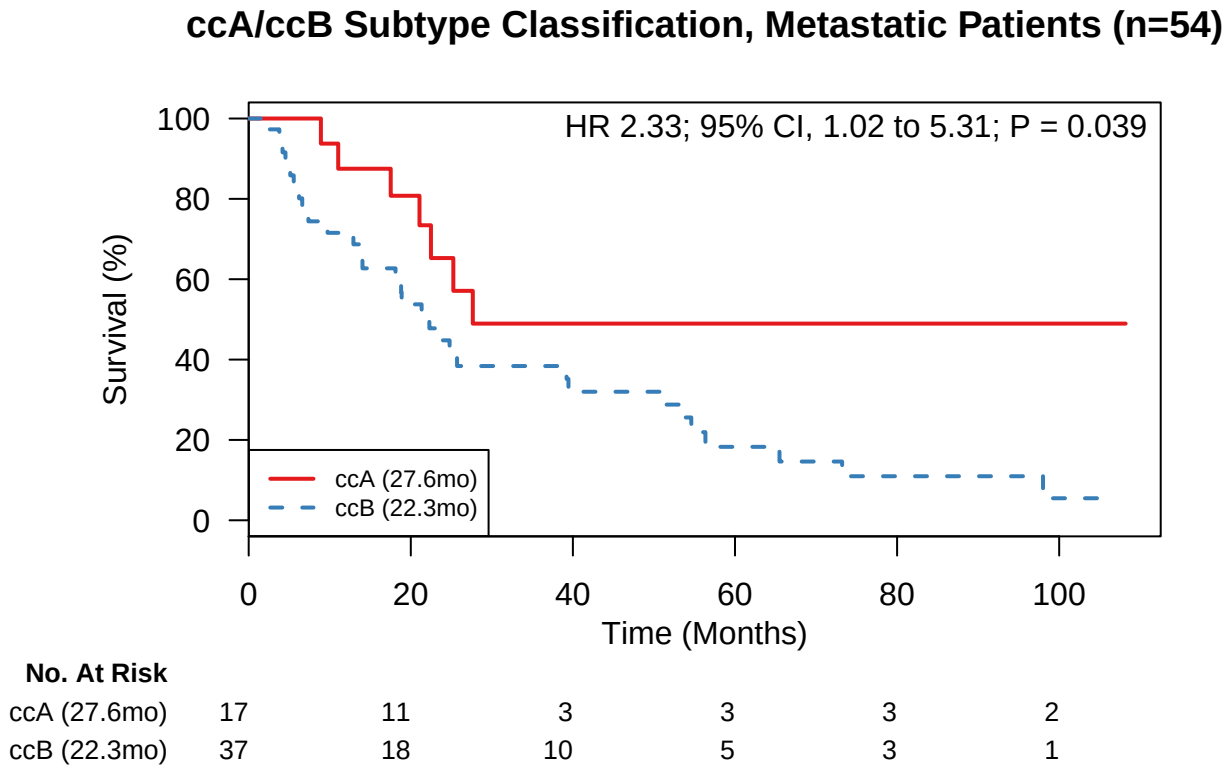
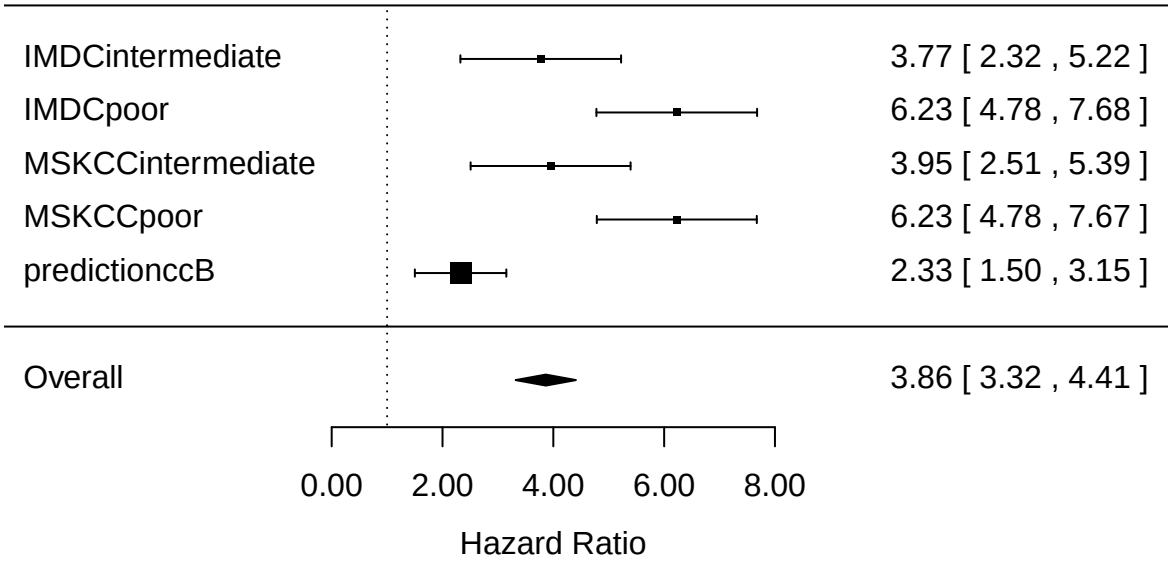


Figure 7: ClearCode34 ccB subtype has worse overall survival in mRCC

Cox PH analysis of IMDC, MSKCC, Clearcode34 (univariate analysis)

In Cox PH univariate analysis the IMDC and MSKCC risk criertia had a greater HR than Clearcode34 subtype, but all three were associated with outcome.

Cox PH univariate models of ClearCode34, IMDC, MSKCC



Concordance Index Analysis -Effect of Tau

The C-Index is interpretable as the probability that a patient predicted to be at lower risk than another patient will survive longer than that patient: its expected value is 0.5 for random predictions, and 1 for a perfect risk model. I used Uno's version of the Concordance Index (or C-Index) using the SurvC1 packages in R. There are different algorithms in R to predict concordance index (see the blog http://gaodoris.blogspot.com/2012/10/5-ways-to-estimate-concordance-index.html for more info)

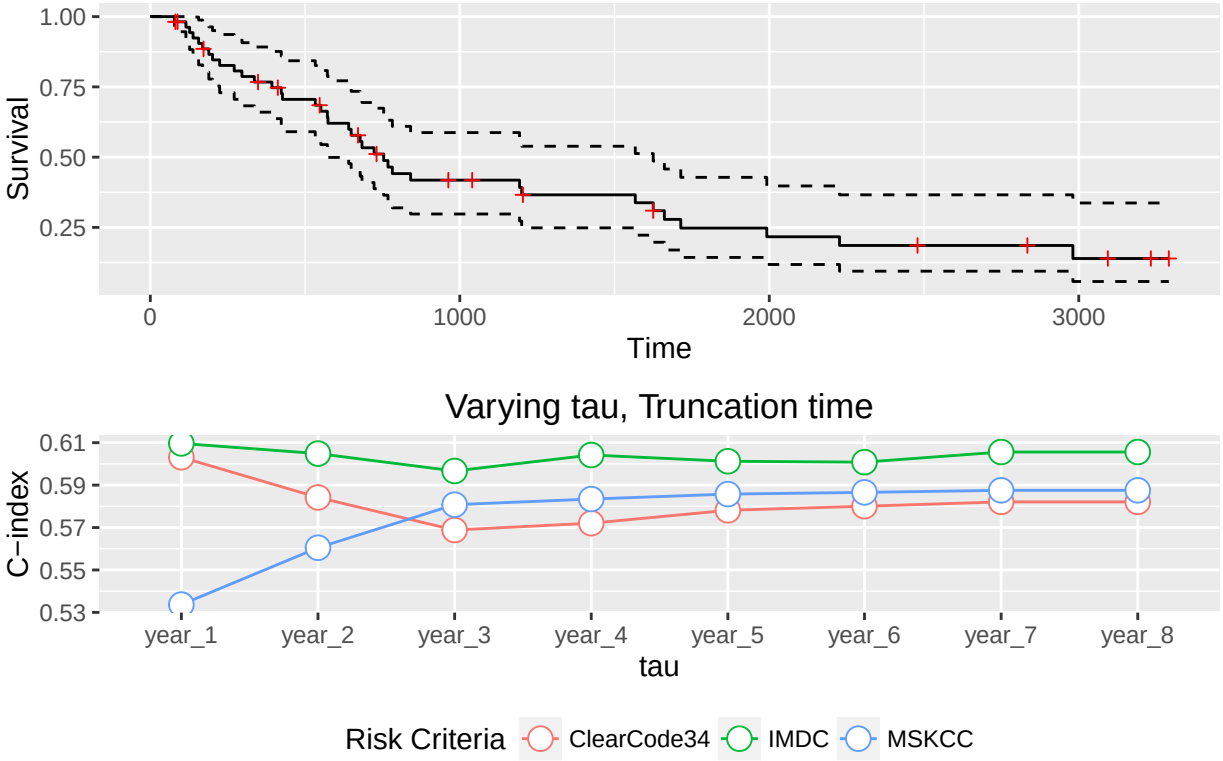
MSKCC and subtype were encoded as dummy variables. Overall Survival was censored (tau) at 3 years.

In our analysis, tau, Truncation time was 3 years. Therefore these results provide a C-index which tells how well the given prediction model works in predicting events that occur in the time range from 0 to tau. Note that the survival function for the underlying censoring time distribution needs to be positive at tau.

```
## ClearCode34        IMDC       MSKCC
##       0.569       0.597       0.581
```

We examined the effect of varying tau and observed that when there are few events that the MSKCC criteria appears to be less robust c-index values, however after 3 years, all risk criteria appears to have a stable c-index

```
##                year_1    year_2    year_3    year_4    year_5    year_6
## ClearCode34 0.6031200 0.5841813 0.5688233 0.5720348 0.5781145 0.5800416
## IMDC        0.6096432 0.6048596 0.5966548 0.6041815 0.6012354 0.6008136
## MSKCC       0.5336813 0.5604733 0.5807668 0.5834233 0.5857619 0.5865699
##                year_7    year_8
## ClearCode34 0.5820786 0.5820786
## IMDC        0.6055979 0.6055979
## MSKCC       0.5875299 0.5875299
```



Concordant with the Cox PH analysis, the c-index of IMDC (0.597) was greater than MSKCC (0.581) which was greater than ClearCode34 (0.569), where 0.5 is random and a c-index >0.5 is associated with worse

outcome.

Consistent with the CoxPH survival analysis, Clearcode34, MSKCC and IMDC are all prognostic in mRCC. However we observed earlier than MSKCC and IMDC have highly significantly overlapping case risk -criteria. By constrast the ccA/ccB classification was discordant with the MSKCC and IMDC risk criertia. Therefore we asked if Clearcode34 provided additional prognostic information in multi-variate models

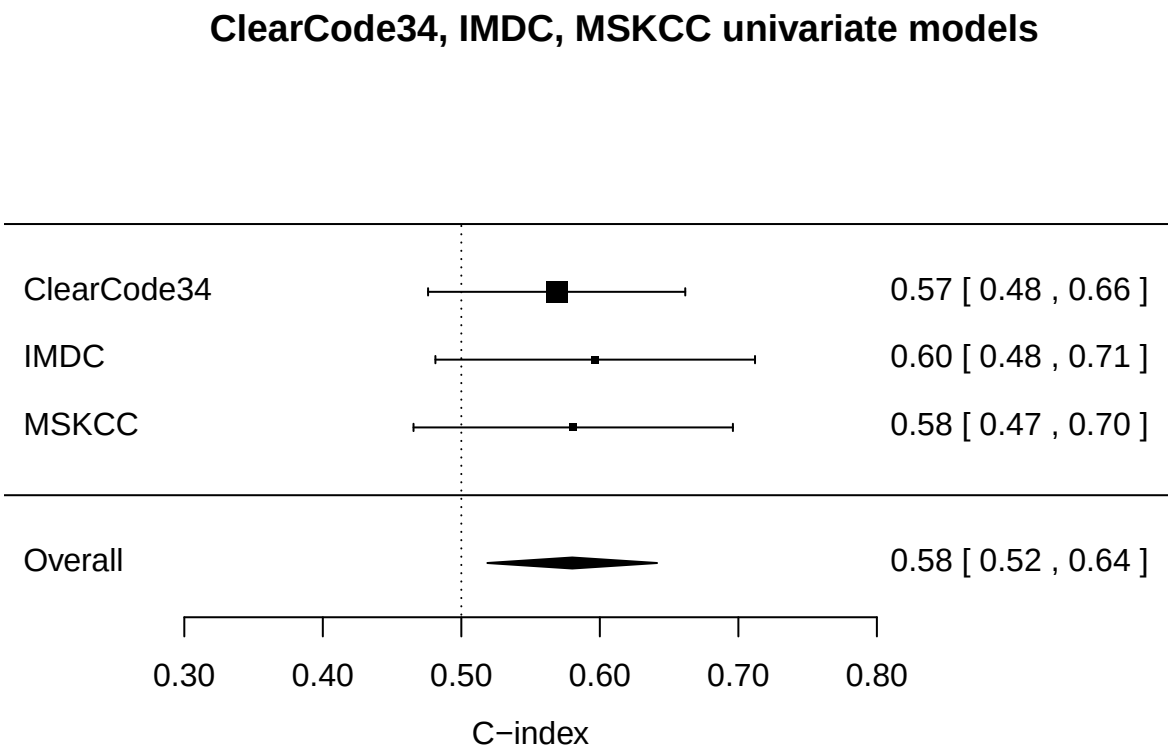


Figure 8:

Comparative performance of Clearcode34, IMDC and MSKCC risk criteria in predicting mRCC survival- Multivariate Analysis

We observed that addition of ClearCode34 ccA/ccB predicted subtypes to the Cox PH survival models of mRCC, significantly improved the fit of the Cox PH model.

Adding the ClearCode34 ccA/ccB subtypes to a model with IMDC was a better fit than just IMDC alone ($p<0.05$)

Simarily, adding the ClearCode34 subtypes also improved the fit of MSKCC risk critertia than just MSKCC alone ($p<0.05$)

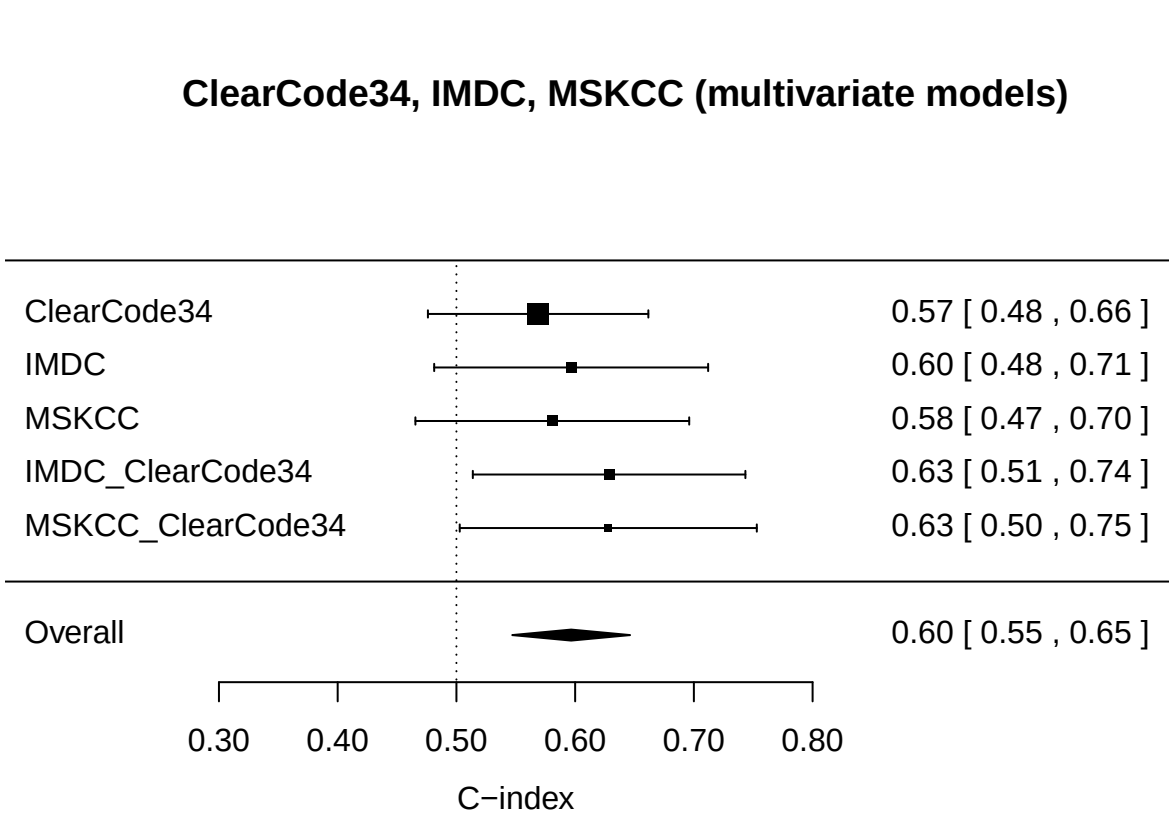
```
## [1] "Testing CoxPH models of MSKCC or ClearCode34 predicted subtype with MSKCC, log rank test"
```

```
## Analysis of Deviance Table
##  Cox model: response is  Surv(TIME, EVENT)
##  Model 1: ~ MSKCC
##  Model 2: ~ prediction + MSKCC
##    loglik  Chisq Df P(>|Chi|)
## 1 -115.96
## 2 -113.48 4.9765  1   0.02569 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## [1] "Testing CoxPH models of IMDC vs ClearCode34 predicted subtype with IMDC, log rank test"
```

```
## Analysis of Deviance Table
##  Cox model: response is  Surv(TIME, EVENT)
##  Model 1: ~ IMDC
##  Model 2: ~ prediction + IMDC
##    loglik  Chisq Df P(>|Chi|)
## 1 -115.64
## 2 -113.63 4.0287  1   0.04473 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Similar results were observed in multivariate concordance index analysis where a model of ClearCode34 and IMDC (or MSKCC) had a higher c-index (0.63, 0.63) compared to any individual univariate model (0.57, 0.60, 0.58, tau= 3 years)



The Choudhury14 8-gene Model: Building and reproducing published results

The coefficients provided in Supplementary Table 7 of Choudhury et al., were used to derive a class-based outcome variable for Cox proportional hazards and concordance index survival analysis.

All data analysis was performed on data that were log2 transformed (previously) since the model was optimized on a qPCR platform (where the data are log2 transformed).

Building the Choudhury14 model

```
##   CXCL5   EFNA5    EMCN   LAMB3     PLG   PRAME RARRES1 SLC6A19
## -0.0182 -0.0364  0.0743 -0.0590  0.0506 -0.1320 -0.0550  0.1050
```

```
## Call:
## coxph(formula = Surv(TIME, EVENT) ~ CXCL5 + EFNA5 + EMCN + LAMB3 +
##     PLG + PRAME + RARRES1 + SLC6A19, data = cbind(pData(TCGAMetc8),
##     t(exprs(TCGAMetc8))), init = coefs, iter.max = 0)
##
##   n= 54, number of events= 37
##    (2 observations deleted due to missingness)
##
##             coef exp(coef) se(coef)      z Pr(>|z|)
## CXCL5   -0.01820   0.98196  0.08651 -0.210    0.833
## EFNA5   -0.03640   0.96425  0.15344 -0.237    0.812
## EMCN     0.07430   1.07713  0.25501  0.291    0.771
## LAMB3   -0.05900   0.94271  0.18668 -0.316    0.752
## PLG      0.05060   1.05190  0.05906  0.857    0.392
```

```
## PRAME   -0.13200   0.87634  0.09408 -1.403    0.161
## RARRES1 -0.05500   0.94649  0.22069 -0.249    0.803
## SLC6A19  0.10500   1.11071  0.08375  1.254    0.210
##
##         exp(coef) exp(-coef) lower .95 upper .95
## CXCL5      0.9820     1.0184    0.8288     1.163
## EFNA5      0.9643     1.0371    0.7138     1.303
## EMCN       1.0771     0.9284    0.6534     1.776
## LAMB3      0.9427     1.0608    0.6538     1.359
## PLG        1.0519     0.9507    0.9369     1.181
## PRAME      0.8763     1.1411    0.7288     1.054
## RARRES1    0.9465     1.0565    0.6141     1.459
## SLC6A19    1.1107     0.9003    0.9426     1.309
##
## Concordance= 0.408  (se = 0.054 )
## Rsquare= 0   (max possible= 0.996 )
## Likelihood ratio test= 0  on 8 df,   p=1
## Wald test            = 0  on 8 df,   p=1
## Score (logrank) test = 84.72  on 8 df,   p=5.44e-15
```

Building and reproducing of Choudhury 8 gene model (reproducing Fig 2B from Choudhury et al.,)

To ensure we had implemented the Choudhury model faithfully, we reproduced an analysis from their article (Fig 2B) in which apply their model to n=419 TCGA tumors.

In Choudhury et al., they report that the 8-gene model identified a large subset of good (n=310) and a smaller subset of poor prognosis (n=103) in TCGA patients (Figure 2B). They reported that the poor prognosis subtype had a hazard ratio of 2.26, 95% CI 1.59 to 3.21, p value 3.04 x 10-6 compared to the favorable subtype.

Unfortunately, they do not provide this list of 419 samples with the article or supplement so I tried to define as close as possible to the 419 patient that they might have used. Firstly, TCGA sample (n=606) were subset to primary Tumors (n=533). Since Choudhury et al., downloaded their data on Oct 2nd 2013, we subseted the 533 tumors to those for which clinical annotation was available in Oct 2013 (n=470 tumors) and were included in the KIRC TCGA publication (Nature, Volume 499 Number 7456, July 4, 2013). This list was downloaded from https://tcga-data.nci.nih.gov/docs/publications/kirc_2013/ and provided a list of 418 samples, which is close to the sample size (n=419) described in the publication.

When we applied the 8-gene model, we identifed two subgroups a larger favorable (n=289) and a smaller group poor (n=129) prognosis patients.

```
##
## favorable      poor
##       289       129
```

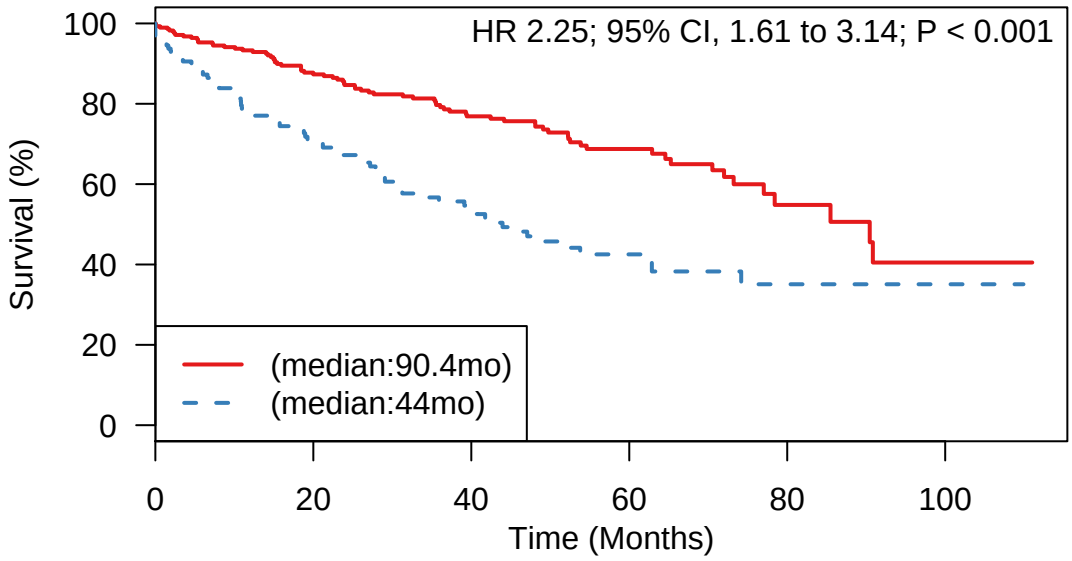
```
## Call:
## coxph(formula = Surv(TIME, EVENT) ~ predictionClass, data = pData(Valdata))
##
##   n= 418, number of events= 140
##
##                       coef exp(coef) se(coef)     z Pr(>|z|)
## predictionClasspoor 0.8110    2.2501   0.1698 4.777 1.78e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
##                       exp(coef) exp(-coef) lower .95 upper .95
## predictionClasspoor        2.25     0.4444     1.613     3.138
##
## Concordance= 0.61  (se = 0.021 )
## Rsquare= 0.051   (max possible= 0.973 )
## Likelihood ratio test= 21.83  on 1 df,   p=2.981e-06
## Wald test            = 22.82  on 1 df,   p=1.776e-06
## Score (logrank) test = 24.09  on 1 df,   p=9.186e-07
```

```
##                             records n.max n.start events   median  0.95LCL
## predictionClass=favorable       289   289     289     75 90.44384 77.03014
## predictionClass=poor            129   129     129     65 43.95616 31.26575
##                              0.95UCL
## predictionClass=favorable         NA
## predictionClass=poor        74.16986
```

```
##                                    25       50 75
## predictionClass=favorable 48.09863 90.44384 NA
## predictionClass=poor      15.74795 43.95616 NA
```

Reproducing Choudhury et al, Fig 2B − TCGA tumors(n=418)



No. At Risk

ıedian:90.4mo)	289	201	130	70	16	3
(median:44mo)	129	78	51	25	6	1

We found that the larger poor prognosis group (n=289) had a hazard ratio of 2.26, 95% CI 1.59 to 3.21, p value 3.04 x 10-6 compared to the good prognosis group (n=129). The two groups had a median survival of 91.7 and 44.6 months.

These results are similar to what was published by Choudhury et al., They reported that the poor prognosis group had a HR of 2.25, with the 95% CI of 1.61 to 3.14 with a log rank test p value of 2.981e-06 compared to the smaller good prognosis groups. They reported that the good and poor subtypes with a median survival per group at 90.4 and 44 months.

Classification of the mRCC cohort (n=56) using the Choudhury 8-gene model

We applied the 8 gene model to the mRCC cases. The data were log2+1 transformed as recommended by Choudhury et al.,

The 8 gene model predicted two groups, a favorable prognosis group (n=25, with a median survival of 22.3 months), and a poor prognosis group (n=31, with a median survival of 27.65 months).

The poor prognosis group was not significantly associated with worse outcome in mRCC (HR of 1.68 with 95% CI of 0.85 to 3.32 and a Log rank test p value of 0.134).

```
##
## favorable      poor
##        25        31
```

```
## ChoudhurypredictionClass=favorable      ChoudhurypredictionClass=poor
##                            27.64932                            22.29041
```

Choudhury 8-gene model and IMDC/MSKCC risk criteria are distinct.

Whilst the MSKCC and IMDC risk criteria are significantly overlapping, there was no significant overlap in these risk criteria and the poor, favorable groups predicted by the Choudhury 8 gene model. For example, tumors predicted to have favourable MSKCC or IMDC risk criteria were equally split between good and poor outcome by the C8 gene model.

Only 4 of the 25 tumor with C8 gene model good prognosis were also predicted to have favorable outcome in the MSKCC or IMDC critera.

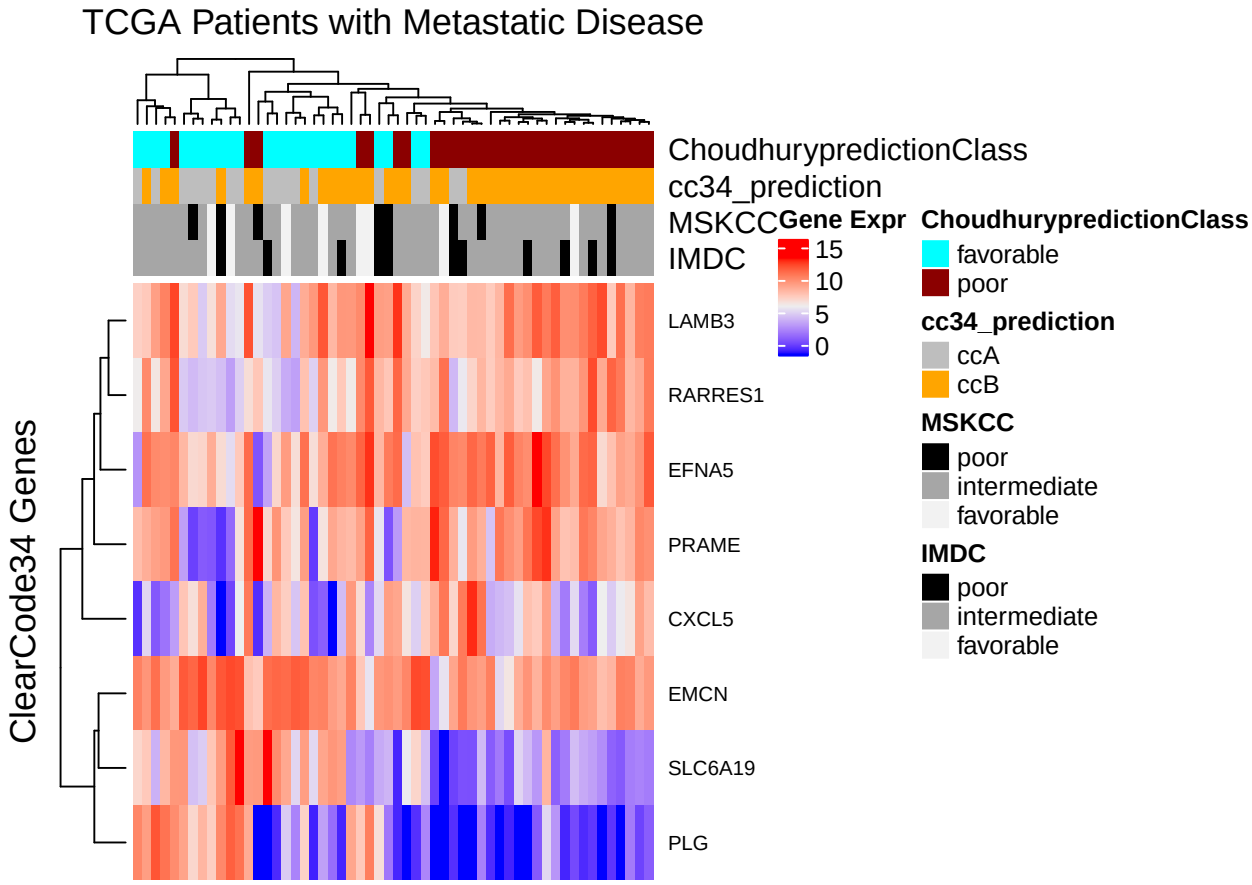
Only 6 or 4 of the 31 tumors with C8 gene model poor prognosis were also predicted to have poor outcome in the IMDC or MSKCC critera respectively.

There was no significant overlap in the Choudhury 8 gene model classification and the IMDC risk (p=0.94) or MSKCC risk groups (p=0.88).

```
## [1] "Overlap with IMDC"

##
##             favorable intermediate poor
##   favorable         4           16    5
##   poor              4           21    6

##
##  Pearson's Chi-squared test
##
## data:  table(TCGAMet$ChoudhurypredictionClass, TCGAMet$IMDC)
## X-squared = 0.12516, df = 2, p-value = 0.9393

## [1] "Overlap with MSKCC"

##
##             favorable intermediate poor
##   favorable         4           17    4
##   poor              4           23    4

##
##  Pearson's Chi-squared test
##
## data:  table(TCGAMet$ChoudhurypredictionClass, TCGAMet$MSKCC)
## X-squared = 0.26013, df = 2, p-value = 0.878
```

Gene Expression of the 8 genes in mRCC (Heatmap)

Data were log2 transformed (log2+1) and median centered (using the row medians of the 533 dataset) as above. Unsupervised hierarchical cluster analysis (distance was 1- Pearson Correlation Coefficient, with average linkage) was applied to the gene expression (RNAseq) profiles of the 8 genes Choudhury14 signature in the 56 mRCC tumor and to the entire n=533 tumors.



The Choudhury 8-gene model does not significantly predict overall survival in mRCC (p=0.134)

There was a trend but not a significant association between the Choudhury 8-gene model subtypes and overall survival in mRCC.

Univariate coxph analysis of Choudhury 8 gene model in mRCC

We applied the 8 gene model to the 56 patient with metastatic disease in the TCGA cohort. The data with log2+1 transformed as recommended by Choudhury et al.,

The 8 gene model predicted two groups, a favorable prognosis group (n=25, with a median survival of 22.3 months), and a poor prognosis group (n=31, with a median survival of 27.6 months).

Although there was a trend, there was no significant different in survival between these two groups (HR was 1.68 with 95% CI of 0.85 to 3.32 and a Log rank test p value of 0.134).

```
## Call:
## coxph(formula = Surv(TIME, EVENT) ~ factor(ChoudhurypredictionClass),
##     data = pData(TCGAMet))
##
##   n= 54, number of events= 37
##    (2 observations deleted due to missingness)
##
##                                       coef exp(coef) se(coef)     z
## factor(ChoudhurypredictionClass)poor 0.5169    1.6769   0.3490 1.481
##                                      Pr(>|z|)
## factor(ChoudhurypredictionClass)poor    0.139
##
##                                      exp(coef) exp(-coef) lower .95
## factor(ChoudhurypredictionClass)poor     1.677     0.5963    0.8462
##                                      upper .95
## factor(ChoudhurypredictionClass)poor     3.323
##
## Concordance= 0.563  (se = 0.047 )
## Rsquare= 0.041   (max possible= 0.988 )
## Likelihood ratio test= 2.29  on 1 df,   p=0.1305
## Wald test            = 2.19  on 1 df,   p=0.1385
## Score (logrank) test = 2.24  on 1 df,   p=0.1345
```

```
## Call: survfit(formula = Surv(TIME, EVENT == 1) ~ ChoudhurypredictionClass,
##     data = pData(TCGAMet))
##
##    2 observations deleted due to missingness
##                                     n events median 0.95LCL 0.95UCL
## ChoudhurypredictionClass=favorable 23     13   27.6    21.1      NA
## ChoudhurypredictionClass=poor      31     24   22.3    14.0    51.5
```

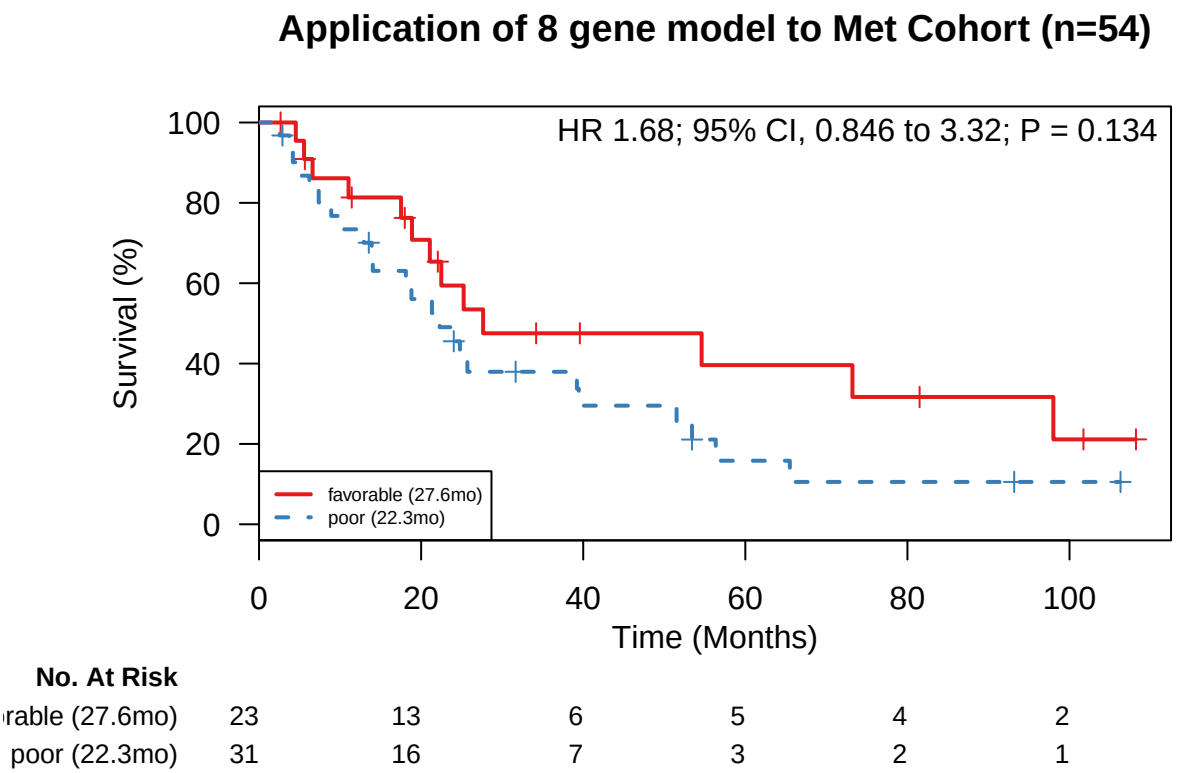


Figure 9:

Multivariate Survival Analysis of 8-gene Choudhury and IMDC or MSKCC risk groups

There was no statistically significant improvement between univariate or multivariate models that included the Choudhury 8-gene model subtypes and IMDC or MSKCC risk groups (p >0.5).

Models were tested using both Cox PH and Concordance Index survival analysis.

When the C8 risk criteria groups were added to the IMDC risk criteria, it did not significant improve the model fit (p = 0.25)

When the C8 risk criteria groups were added to the IMDC risk criteria, it did not significant improve the model fit (p = 0.25)

```
## [1] "Anova coxph of IMDC vs subtype + IMDC, log rank test"
```

```
## Analysis of Deviance Table
##  Cox model: response is  Surv(TIME, EVENT)
##  Model 1: ~ factor(IMDC)
##  Model 2: ~ factor(ChoudhurypredictionClass) + factor(IMDC)
##    loglik  Chisq Df P(>|Chi|)
## 1 -115.64
## 2 -114.99 1.3018  1    0.2539
```

When the C8 risk criteria groups were added to the MSKCC risk criteria, it did not significant improve the model fit (p = 0.29)

```
## [1] "Anova coxph of MSKCC vs Choudhury Risk + MSKCC, log rank test"
```

```
## Analysis of Deviance Table
```

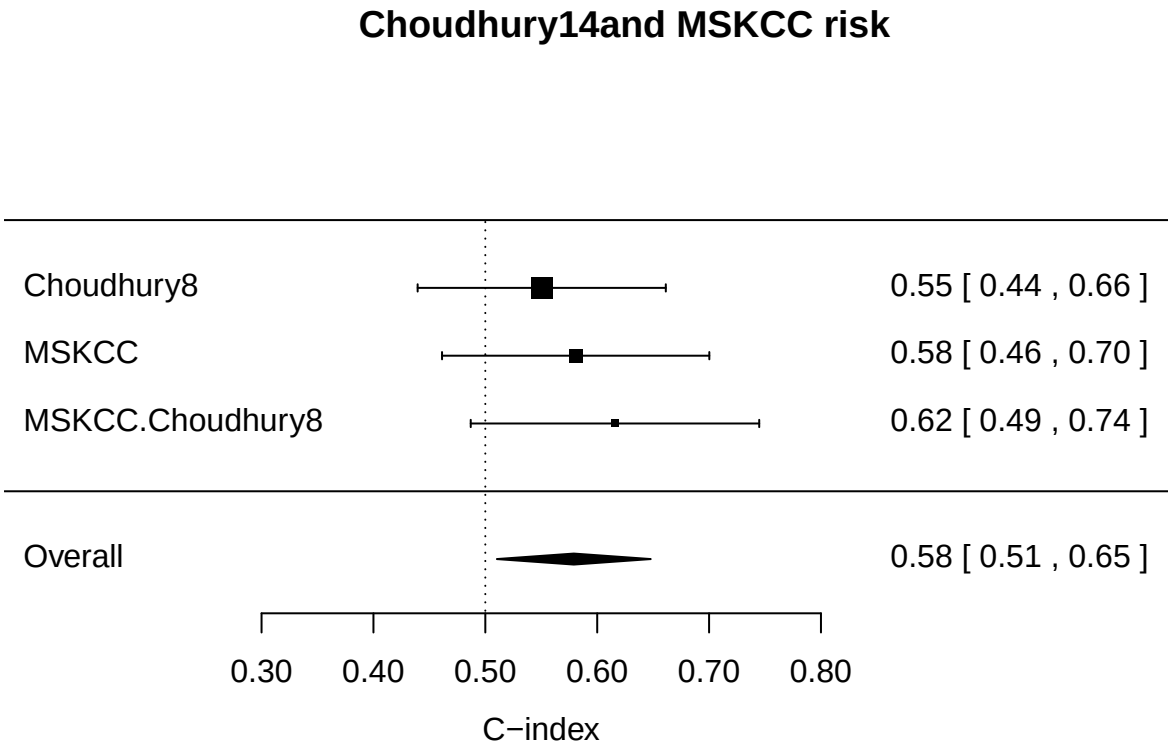
```
##  Cox model: response is  Surv(TIME, EVENT)
##  Model 1: ~ factor(MSKCC)
##  Model 2: ~ factor(ChoudhurypredictionClass) + factor(MSKCC)
##    loglik  Chisq Df P(>|Chi|)
## 1 -115.96
## 2 -115.40 1.1313  1    0.2875
```

In Concordance Index analysis of MSKCC, IMDC and Choudhury 8 gene subtype, IMDC,MSKCC and risk subtype were encoded as dummy variables and tau=3 years was used as described above

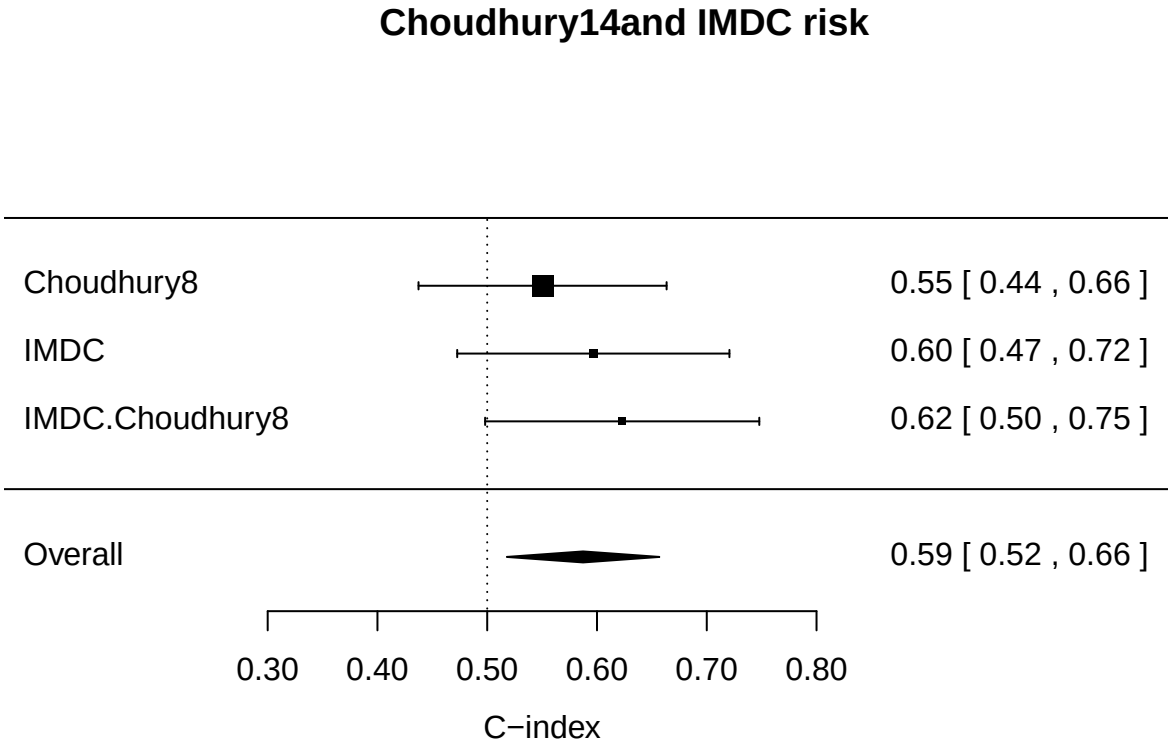
The Choudhury 8-gene model has a c-index of 0.55 which was lower than the Clearcode34 model (0.57), the IMDC risk criteria (0.60) or MSKCC risk criteria (0.58).

In multi-variate model, there was trend towards higher c-index in the joint models; MSKCC + Choudhury8 (0.62), IMDC.Choudhury8 (0.62), however the model Delta was not significantly different.

```
##                          Est         SE    Lower95   Upper95
## MSKCC.Choudhury8 0.61588158 0.05879321  0.5007057 0.7310575
## Choudhury8       0.55041415 0.05658494  0.4395642 0.6612641
## Delta            0.06546743 0.06813156 -0.0680023 0.1989372
```




```
##                   Est    SE Lower95 Upper95
## IMDC.Choudhury8 0.623 0.066   0.494   0.752
## Choudhury8      0.550 0.058   0.437   0.663
## Delta           0.073 0.079  -0.083   0.228
```



Forest plot of concordance index results (univariate and multivariate)

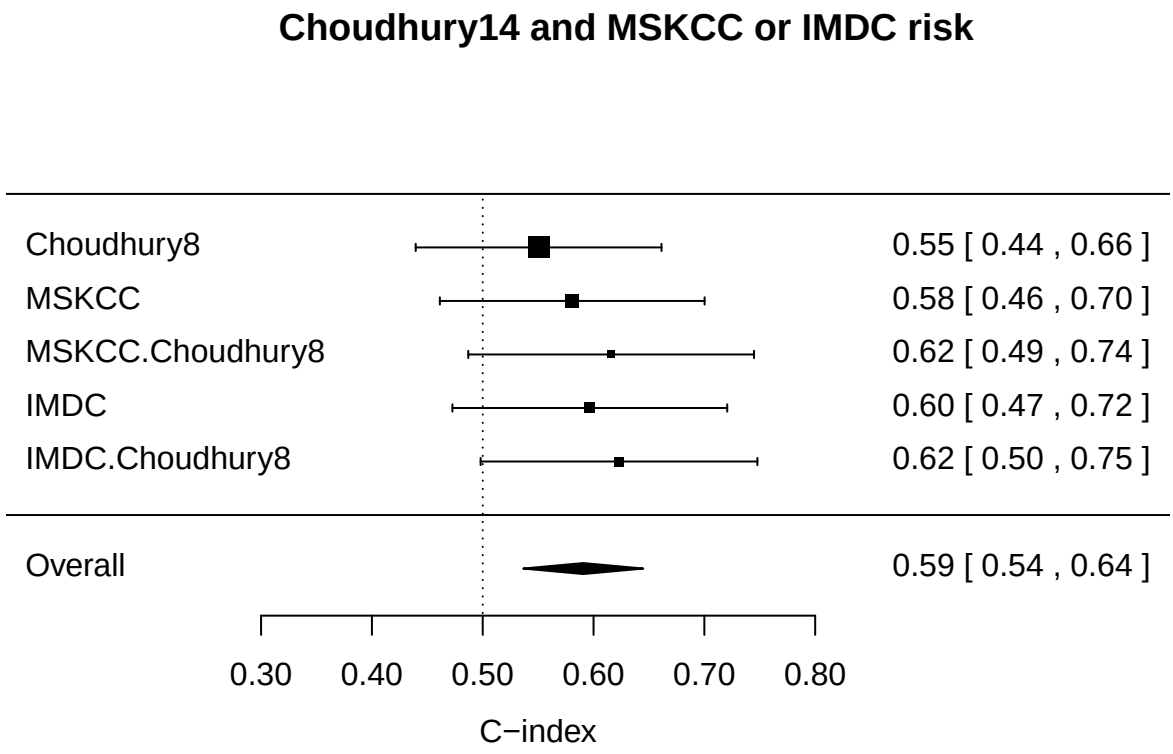


Figure 10:

Given that the ClearCode34 ccA/ccB subtypes were significantly different, but the Choudhury was weakly

but not significantly prognostic, we compared these gene signatures in detail

Detailed Comparision of Clearcode34 and Choudhury 8 gene signatures

```
## [1] "/Users/aedin"
```

```
## [1] TRUE
```

```
## [1] TRUE
```

No overlap in genes in ClearCode34 and Choudhury 8 gene signature

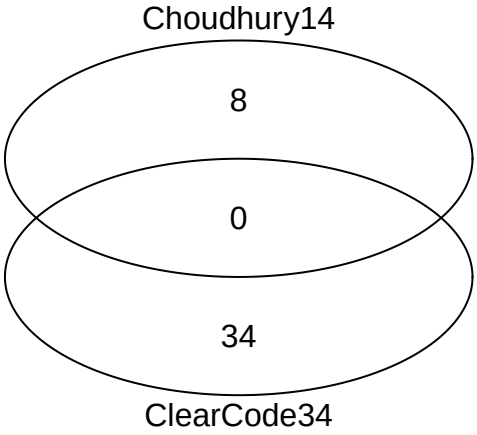
The clearcode34 genes are

MAPT, STK32B, FZD1, RGS5, GIPC2, PDGFD, EPAS1, MAOB, CDH5, TCEA3, LEPROTL1, BNIP3L, EHBP1, VCAM1, PHYH, PRKAA2, SLC4A4, ESD, TLR3, NRP1, C11ORF1, ST13, ARNT, SPRYD7, SERPINA3, SLC4A3, MOXD1, KCNN4, ROR2, FLJ23867, FOXM1, CCNO, GALNT10, GALNT4

and the genes in the Choudhury 8 gene signature are

CXCL5, EFNA5, EMCN, LAMB3, PLG, PRAME, RARRES1, SLC6A19

There is no overlap in gene signatures.



No significant overlap in genesets or pathways enriched in the ClearCode34 and Chodhury 8 gene signature

We compared both genes signature at the level of gene set and pathway enrichment.

Using DOSE and reactomePA R package we ran gene set enrichment analysis of the Clearcode 34 and the Choudhury et al., 8 gene signature against - the reactome database of 1526 pathways - the biological process (BP) subsets of gene ontology - the molecular function (MF) subsets of gene ontology - the cellular component (CC) subsets of gene ontology,

Using the Bioconductor/R libraries reactomePA (function enrichPathway) and clusterprofiler (enrichGO). P values were adjusted for multiple testing by controlling the False Discovery rate (fdr), also called the Benjamini & Hochberg (1995) correction. Only gene sets or pathways with minimum size (minGSsize) of 4 and maximum size (maxGSsize) of 500 (default setting) were studied. (The default min gene set size is 10 and this was modified given the small number of genes in the ClearCode and C8 signatures)

Significant Pathways or Gene Ontology Terms (adjusted p values) associated with each gene set

The C8 or ClearCode 34 signatures were not significantly enriched in any Reactome Pathway, GO molecular function (MF) or cellular compartment (CC) or Biological process (BP) genesets at p.adjust p<0.001.

At p<0.01, the ClearCode34 signature, 2 genes in the gene signature (MAOB/VCAM1) overlapped also occurred in the GO MF term GO:0008131 primary amine oxidase activity and this overlap was significant (p<0.01)

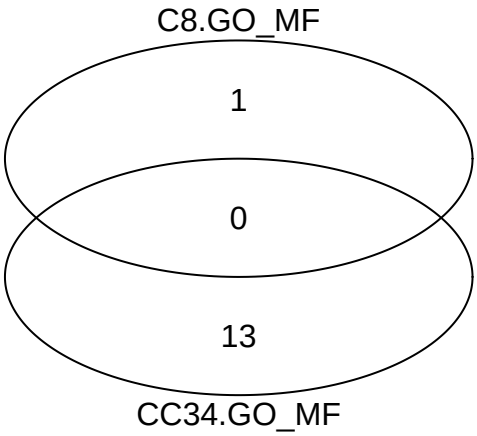
	Pval0.05	Pval0.01	Pval0.001	Pval1e-04
CC34.Pathway	2	0	0	0
C8.Pathway	0	0	0	0
CC34.GO_BP	0	0	0	0
C8.GO_BP	0	0	0	0
CC34.GO_MF	13	1	0	0
C8.GO_MF	21	0	0	0
CC34.GO_CC	0	0	0	0
C8.GO_CC	0	0	0	0

Many of the significant genesets, were because of only 1 gene overlap, which is unlikely to be biologically informative. Therefore, we examined the number of genes that overlapped between the gene signatures and the pathway. The following tables, showed the number of gene signature were a p value (adjusted) of <0.05 or <0.01 and an overlap of 2 or more, or 3 or more genes.

	Pval0.05_n>1	Pval0.01_n>1	Pval0.05_n>2	Pval0.01_n>2
CC34.Pathway	2	0	0	0
C8.Pathway	0	0	0	0
CC34.GO_BP	0	0	0	0
C8.GO_BP	0	0	0	0
CC34.GO_MF	13	1	2	0
C8.GO_MF	1	0	0	0
CC34.GO_CC	0	0	0	0
C8.GO_CC	0	0	0	0

At p.adjust <0.05 and min gene overlap of 2 genes, there were 2 reactome pathways enriched in the ClearCode34 genes and there were 1 and 13 molecular function Gene Ontology (GO-BP) terms enriched in the ccA/ccB and C8 genes signature respectively

Note the ccA/ccB and 8-gene signature were not enriched in similar genesets or pathways.



Pathways and GO terms enriched in the C8 Signature

The C8 gene signature was only enriched ($p<0.05$, with 2 or more genes) in molecualr function cytokine receptor binding (GO:0005126)

```
##                    ID               Description GeneRatio   BgRatio
## GO:0005126 GO:0005126 cytokine receptor binding       2/6 272/16742
##                  pvalue   p.adjust     qvalue      geneID Count
## GO:0005126 0.003778309 0.03320017 0.01350246 CXCL5/EFNA5     2
```

Pathways and GO terms enriched in the ClearCode Signature

The most significant Reactome pathways enriched in the ClearCode34 signature were "Regulation of gene expression by Hypoxia-inducible Factor (HIF)" and "Bicarbonate transporters"

```
##              ID                                                  Description
## 1234158 1234158 Regulation of gene expression by Hypoxia-inducible Factor
## 425381   425381                                 Bicarbonate transporters
##         GeneRatio BgRatio       pvalue   p.adjust      qvalue
## 1234158      2/19  9/6749 0.0002671805 0.01322544 0.009843492
## 425381       2/19  9/6749 0.0002671805 0.01322544 0.009843492
##                geneID Count
## 1234158    EPAS1/ARNT     2
## 425381  SLC4A4/SLC4A3     2
```

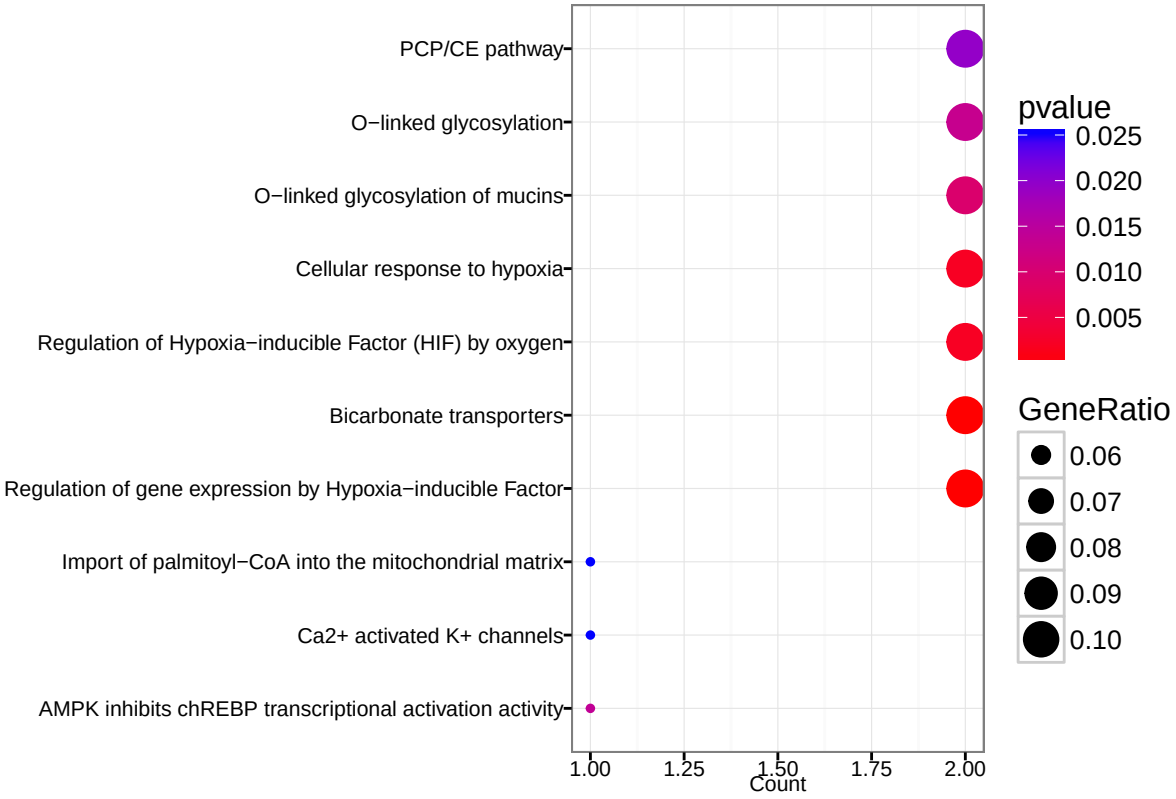
Hypoxia Pathways enriched in the ClearCode34 Signature

The Reactome pathway "Regulation of gene expression by Hypoxia-inducible Factor (HIF)" has 10 genes (ARNT, EP300, HIF1A, CREBBP, EPAS1, CITED2, HIF3A, EPO, VEGFA, CA9) and of these, 2 genes EPAS1 and ARNT were present in the ClearCode34 signatures.

In regulation of gene expression by HIF. HIF-alpha (HIF1A, HIF2A (EPAS1), HIF3A) forms a heterodimer with ARNT (HIF1-beta). Endothelial PAS domain-containing protein 1 (EPAS1, also known as hypoxia-inducible factor-2alpha (HIF-2alpha)) interacts with ARNT.

In this plot below, count is the number of genes that overlapped (1 or 2). Gene Ratio is the count divided by the gene set size. The pathway "Regulation of gene expression by Hypoxia-inducible Factor" was most significant, but several other related pathways were also significant or marginally significant ("Regulation of

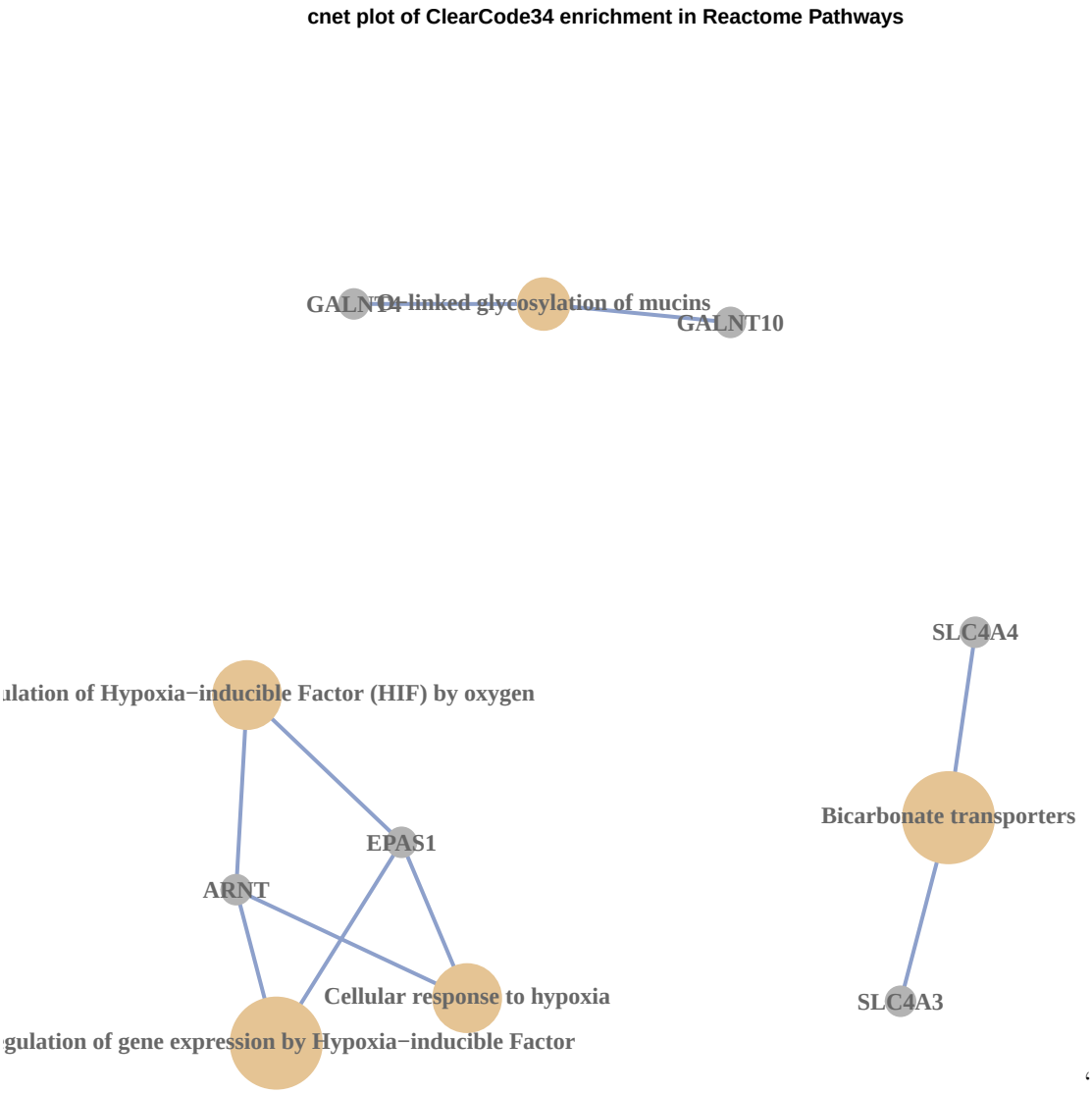
Hypoxia-inducible Factor (HIF) by oxygen", "Cellular response to hypoxia" were ranked 3 and 4). The color shows the unadjusted pvalues.



The network plot belows shows the actual genes that overlapped and the reactome terms that overlapped. The size of each brown cicle indicates the Count (number of genes that overlapped).

The genes EPAS1 and ARNT were associated with Hypoxia related pathway.

The GALNT10 and GALNT4 were enriched in 0-linked glycoylation pathways and the Solute Carrier Family 4 (Anion Exchanger) genes SLC4A3 /SLC4A4 are transporter genes involved in biocarbonate transport.



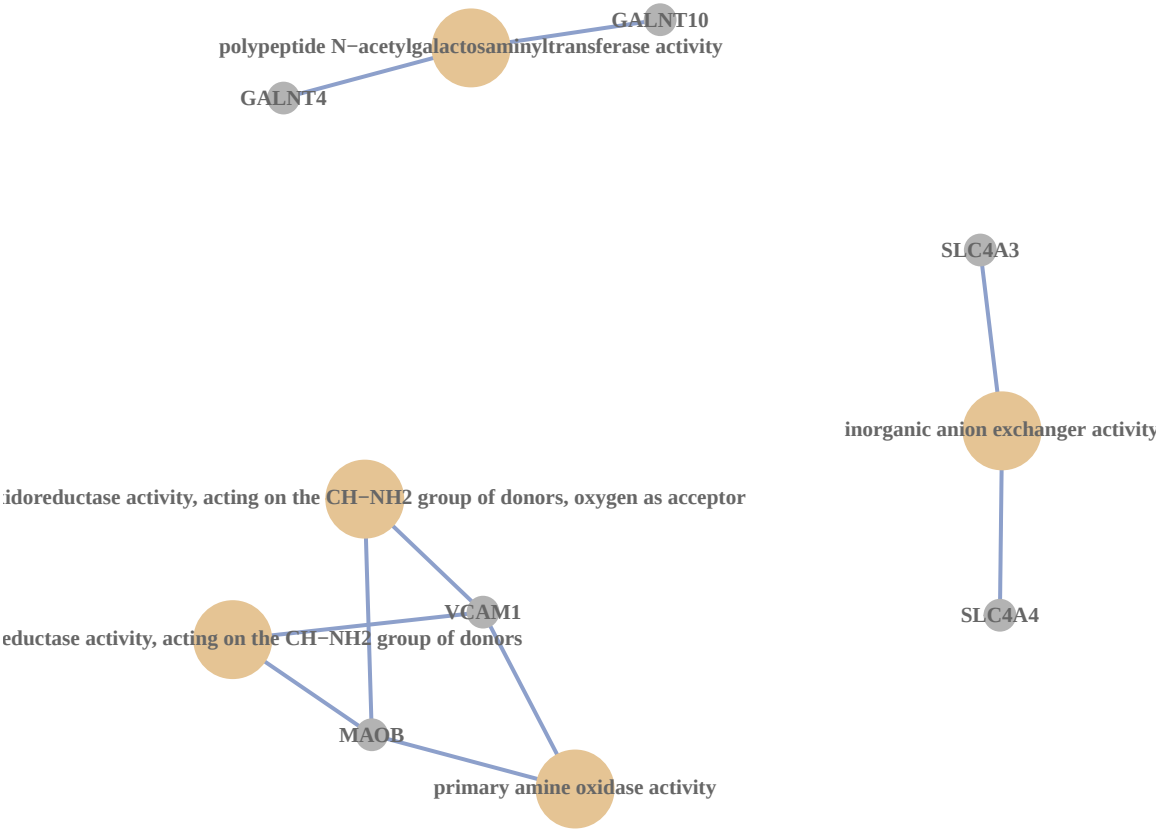
The ClearCode34 signature was enriched in 2 GO molecular functions at p<0.05 with 3 or more gene overlap. The genes MAPT, CDH5, KCNN4 were associated with the molecular function of protein phosphatase binding (GO:0019903) and phosphatase binding (GO:0019902)

```
##                    ID                 Description GeneRatio   BgRatio
## GO:0019903 GO:0019903 protein phosphatase binding      3/31 120/16742
## GO:0019902 GO:0019902         phosphatase binding      3/31 162/16742
##                 pvalue   p.adjust    qvalue          geneID Count
## GO:0019903 0.001394272 0.02451692 0.0170048 MAPT/CDH5/KCNN4     3
## GO:0019902 0.003276589 0.03522333 0.0244307 MAPT/CDH5/KCNN4     3
```

In addition, The ClearCode34 signature was enriched in 13 GO molecular functions at p<0.05 with 2 or more gene overlap. The genes MAOB, VCAM1 were involved in primary amine oxidase activity (GO:0008131) and related GO terms. The genes FZD1, ROR2 were enriched in Wnt-protein binding (GO:0017147) and frizzled binding (GO:0005109)

	ID	Description	GeneRa
GO:0008131	GO:0008131	primary amine oxidase activity	2/31
GO:0016641	GO:0016641	oxidoreductase activity, acting on the CH-NH2 group of donors, oxygen as acceptor	2/31

	ID	Description	GeneRa
GO:0005452	GO:0005452	inorganic anion exchanger activity	2/31
GO:0016638	GO:0016638	oxidoreductase activity, acting on the CH-NH2 group of donors	2/31
GO:0004653	GO:0004653	polypeptide N-acetylgalactosaminyltransferase activity	2/31
GO:0019903	GO:0019903	protein phosphatase binding	3/31
GO:0017147	GO:0017147	Wnt-protein binding	2/31
GO:0015301	GO:0015301	anion:anion antiporter activity	2/31
GO:0015026	GO:0015026	coreceptor activity	2/31
GO:0008376	GO:0008376	acetylgalactosaminyltransferase activity	2/31
GO:0005109	GO:0005109	frizzled binding	2/31
GO:0019902	GO:0019902	phosphatase binding	3/31
GO:0099516	GO:0099516	ion antiporter activity	2/31



Prognostic value and Differntial Gene Expression of enriched pathways in ccA/ccB subtypes in mRCC

Gene in the Choudury signature were only enriched in 1 molecular function but the terms were quite generic and therefore unlikely to be biologically informative and were not explored further.

Genes in ClearCode34 were significantly enriched in pathways or molecular function GO terms associated with hypoxia and HIFvpathway which might be of interest.

The two genes EPAS1/ARNT were in the pathways “Regulation of Hypoxia-inducible Factor (HIF) by oxygen”, “Cellualr response to hypoxia”, ‘Regulation of gene expression by Hypoxia-inducible Factor’, etc. The other genes are given below.

```
##              ID                                                  Description
## 1234158 1234158 Regulation of gene expression by Hypoxia-inducible Factor
## 425381   425381                                Bicarbonate transporters
## 1234174 1234174    Regulation of Hypoxia-inducible Factor (HIF) by oxygen
##         GeneRatio BgRatio       pvalue   p.adjust      qvalue
## 1234158      2/19  9/6749 0.0002671805 0.01322544 0.009843492
## 425381       2/19  9/6749 0.0002671805 0.01322544 0.009843492
## 1234174      2/19 25/6749 0.0021674973 0.05364556 0.039927582
##                geneID Count
## 1234158    EPAS1/ARNT     2
## 425381  SLC4A4/SLC4A3     2
## 1234174    EPAS1/ARNT     2
```

There I tested if all reactome pathways (n=47) to identify pathways which were

1) differentially expressed between ccA and ccB
2) Prognostics in mRCC

First,I performed a differential gene expression analysis of all genes in the dataset (n=20531, but 18502 mapped to pathways) using Limma using moderated variance. Adjusted p values for each gene were combined using the Fisher's combined probability test (using survcomp::: combine.test)

Then I performed a coxph analysis of each set of genes in which the gene set scores was the weighted gene expression sum using the plus-minus model described by Waldron et al.,

```
## [1] "Activation of IRF3/IRF7 mediated by TBK1/IKK epsilon"
## [1] "Activation of PPARGC1A (PGC-1alpha) by phosphorylation"
## [1] "Adherens junctions interactions"
## [1] "AMPK inhibits chREBP transcriptional activation activity"
## [1] "Apoptosis"
## [1] "Apoptotic cleavage of cellular proteins"
## [1] "Apoptotic execution  phase"
## [1] "Bicarbonate transporters"
## [1] "Caspase activation via extrinsic apoptotic signalig pathway"
## [1] "Caspase-mediated cleavage of cytoskeletal proteins"
## [1] "Cellular response to hypoxia"
## [1] "CHL1 interactions"
## [1] "CRMPs in Sema3A signaling"
## [1] "Cyclin A/B1 associated events during G2/M transition"
## [1] "Diseases associated with the TLR signaling cascade"
## [1] "Diseases of Immune System"
## [1] "Energy dependent regulation of mTOR by LKB1-AMPK"
## [1] "IKK complex recruitment mediated by RIP1"
## [1] "Import of palmitoyl-CoA into the mitochondrial matrix"
## [1] "Ligand-dependent caspase activation"
## [1] "Mitochondrial biogenesis"
## [1] "mTOR signalling"
## [1] "O-linked glycosylation"
## [1] "O-linked glycosylation of mucins"
## [1] "Oxygen-dependent proline hydroxylation of Hypoxia-inducible Factor Alpha"
## [1] "PCP/CE pathway"
## [1] "Peroxisomal lipid metabolism"
## [1] "PKB-mediated events"
## [1] "Polo-like kinase mediated events"
## [1] "Programmed Cell Death"
## [1] "Regulation of gene expression by Hypoxia-inducible Factor"
```

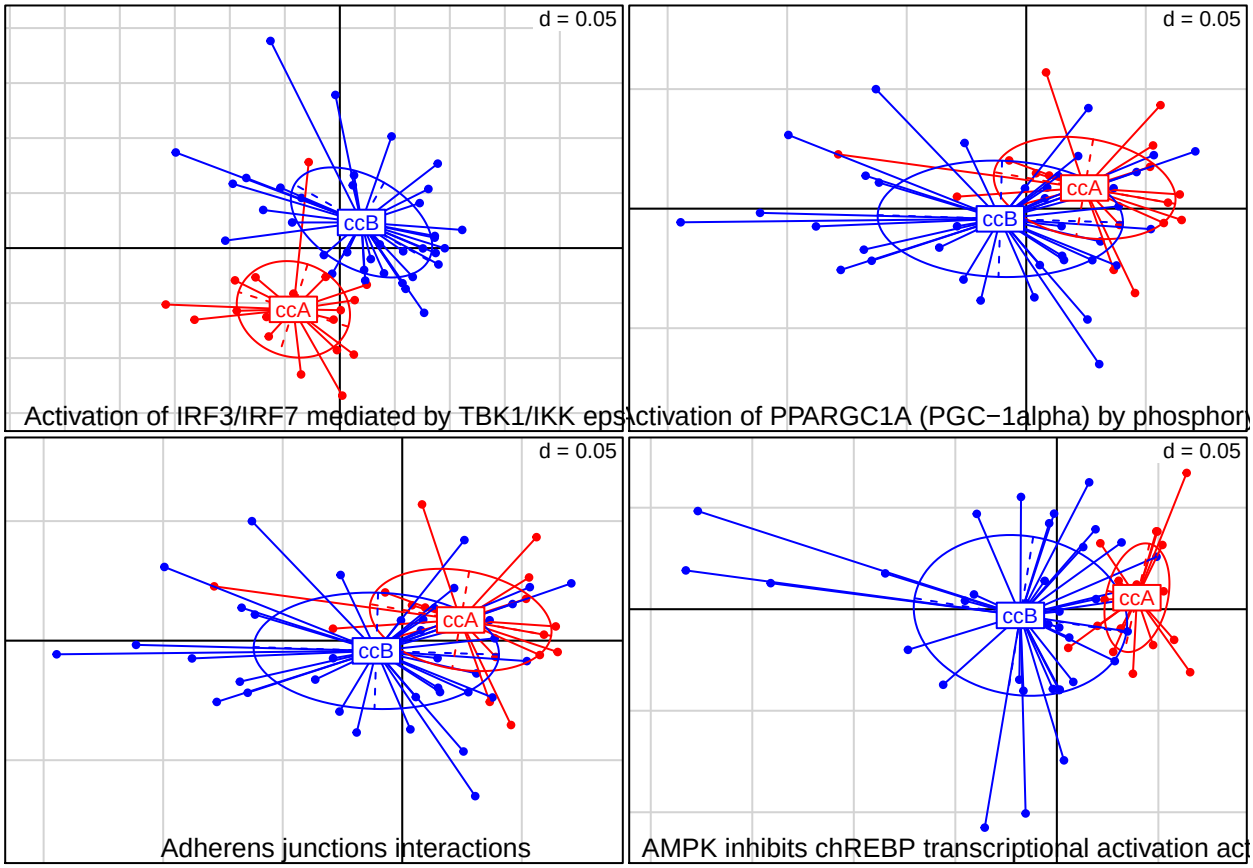
```
## [1] "Regulation of Hypoxia-inducible Factor (HIF) by oxygen"
## [1] "Sema3A PAK dependent Axon repulsion"
## [1] "SEMA3A-Plexin repulsion signaling by inhibiting Integrin adhesion"
## [1] "Signal transduction by L1"
## [1] "Signaling by VEGF"
## [1] "TRAF6 mediated induction of TAK1 complex"
## [1] "Trafficking and processing of endosomal TLR"
## [1] "Transcriptional regulation of pluripotent stem cells"
## [1] "Transport of inorganic cations/anions and amino acids/oligopeptides"
## [1] "TRIF-mediated programmed cell death"
## [1] "VEGFR2 mediated vascular permeability"
```

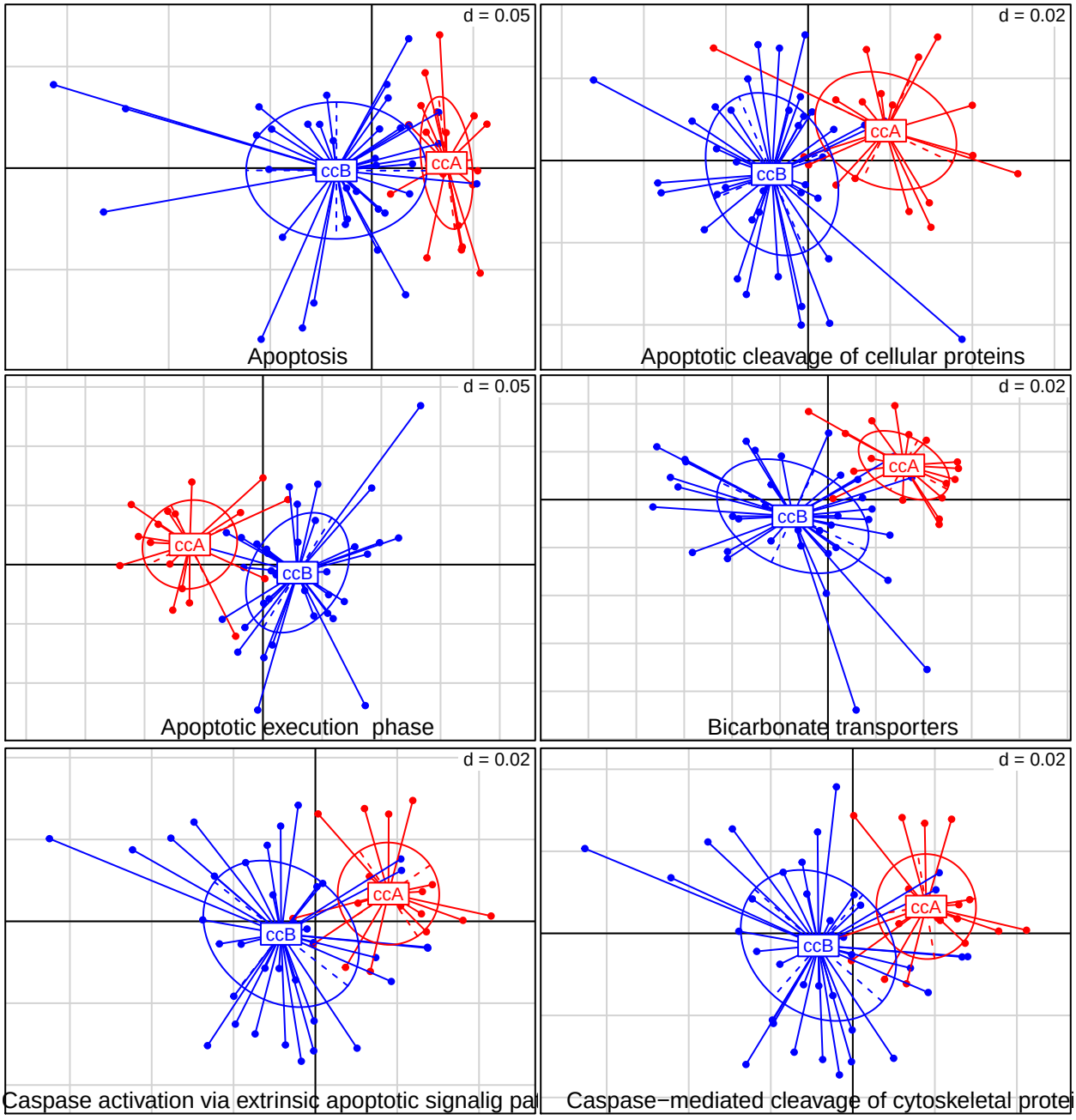
The following 10 pathways were significantly differentially expressed between ccA and ccB (FisherScore adjusted p value of <0.05), prognostic (cox ph adjusted p value <0.05) and enriched with 2 or more gene overlap in the ClearCode34 signature.

```
##
## ------------------------------------------------------------------------------
##                             FisherScore   surv_estimate   surv_p.value
## -------------------------------- ------------- --------------- --------------
## **Bicarbonate transporters**     5.278e-05     0.5463          0.002403
##
## **Cellular response to           0.0101        0.4662          0.008694
## hypoxia**
##
## **Regulation of                  0.0101        0.4662          0.008694
## Hypoxia-inducible Factor (HIF)
## by oxygen**
##
## **O-linked glycosylation of      1.8e-08       0.7857          0.0006042
## mucins**
##
## **O-linked glycosylation**       1.169e-12     0.8132          0.0001183
##
## **PCP/CE pathway**               1.239e-09     0.9924          1.682e-05
##
## **Transport of inorganic         2.791e-12     0.719           0.0003485
## cations/anions and amino
## acids/oligopeptides**
##
## **Signaling by VEGF**            1.525e-14     0.8539          2.061e-06
##
## **Apoptosis**                    1.667e-17     0.8351          1.174e-05
##
## **Programmed Cell Death**        1.667e-17     0.8611          1.174e-05
## ------------------------------------------------------------------------------
##
## Table: Table continues below
##
##
## ---------------------------------------------------------------------
##                             p.adjust   qvalue   geneID
## -------------------------------- ---------- -------- --------------
## **Bicarbonate transporters**     0.01323    0.009843 SLC4A4/SLC4A3
##
```

```
## **Cellular response to           0.05365    0.03993  EPAS1/ARNT
## hypoxia**
##
## **Regulation of                  0.05365    0.03993  EPAS1/ARNT
## Hypoxia-inducible Factor (HIF)
## by oxygen**
##
## **O-linked glycosylation of      0.1656     0.1233   GALNT10/GALNT4
## mucins**
##
## **O-linked glycosylation**       0.1656     0.1233   GALNT10/GALNT4
##
## **PCP/CE pathway**               0.1656     0.1233   FZD1/ROR2
##
## **Transport of inorganic         0.1656     0.1233   SLC4A4/SLC4A3
## cations/anions and amino
## acids/oligopeptides**
##
## **Signaling by VEGF**            0.1656     0.1233   CDH5/NRP1
##
## **Apoptosis**                    0.1699     0.1265   MAPT/TLR3
##
## **Programmed Cell Death**        0.1699     0.1265   MAPT/TLR3
## ---------------------------------------------------------------------
```

Using the genes in the top 10 Reactome pathways, I performed a correspondence analysis to visualize if the gene expression in these genesets distiguished ccA and ccB





Next, I visualize the survival analysis

Significant overlap in mRCC patients classified using the 8 gene score or ClearCode34 ccA/ccB classifiers

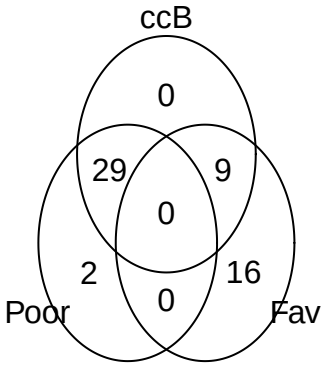
Whilst there was no overlap in genes or pathways in the Clearcode34 or 8 gene signatures, 80% of patients had similar risk classifications andthere was an overlap in patients classifications (p<0.0001). The favorable Choudhury class overlaps with the good prognosis ccA class (16/18) and the poor Choudhury class overlaps with the poorer prognosis ccB class (29/38). However 20% of patients (11/56) has a different classification.

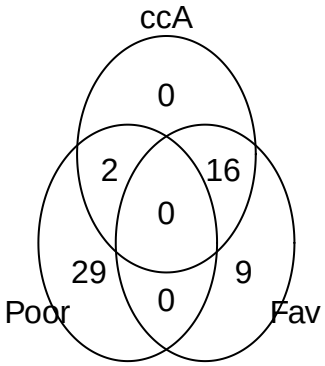
Whilst these criteria have 80% overlap in tumors, (p = 1.737e-05), this overlap is not as significantly overlap between the IMDC and MSKCC critera which have 84% overlap in the 3 risk classes (p-value = 1.635e-13)

```
## [1] "Overlap with MSKCC"
```

```
##
##             ccA ccB
##   favorable  16   9
##   poor        2  29
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(TCGAMet$ChoudhurypredictionClass, TCGAMet$cc34_prediction)
## X-squared = 18.458, df = 1, p-value = 1.737e-05
```





Overlap Choudhury 8 gene and ccB

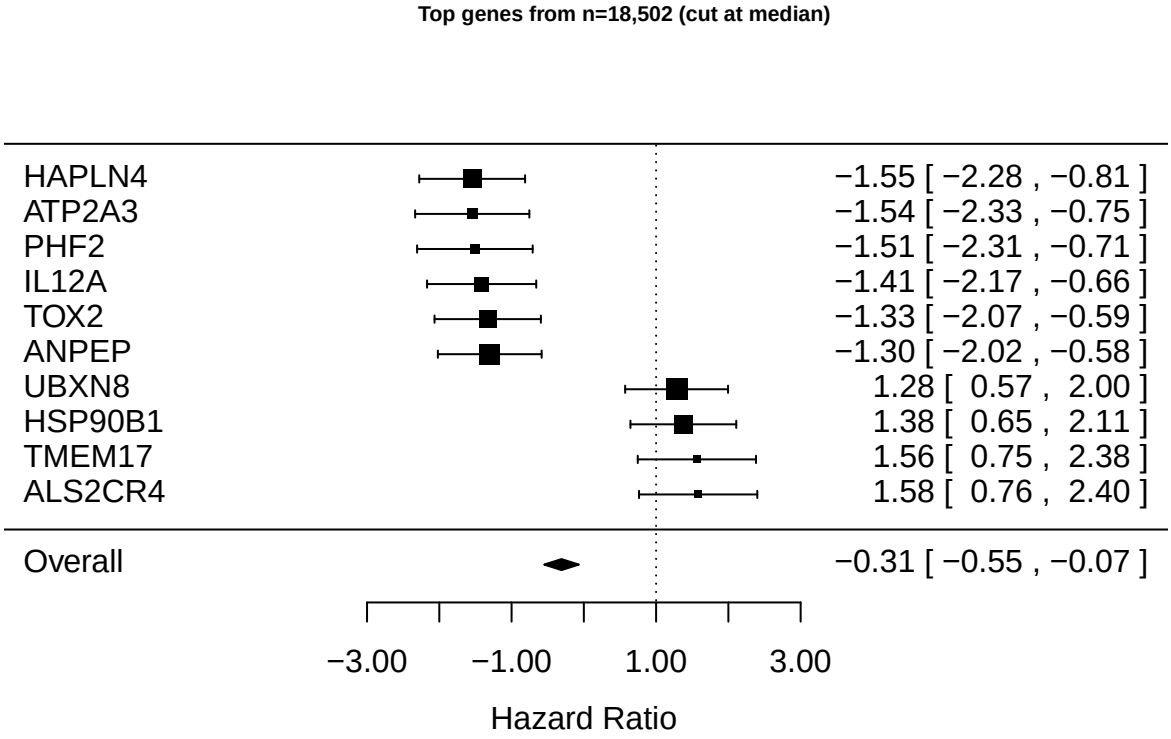
Overlap Choudhury 8 gene and ccA

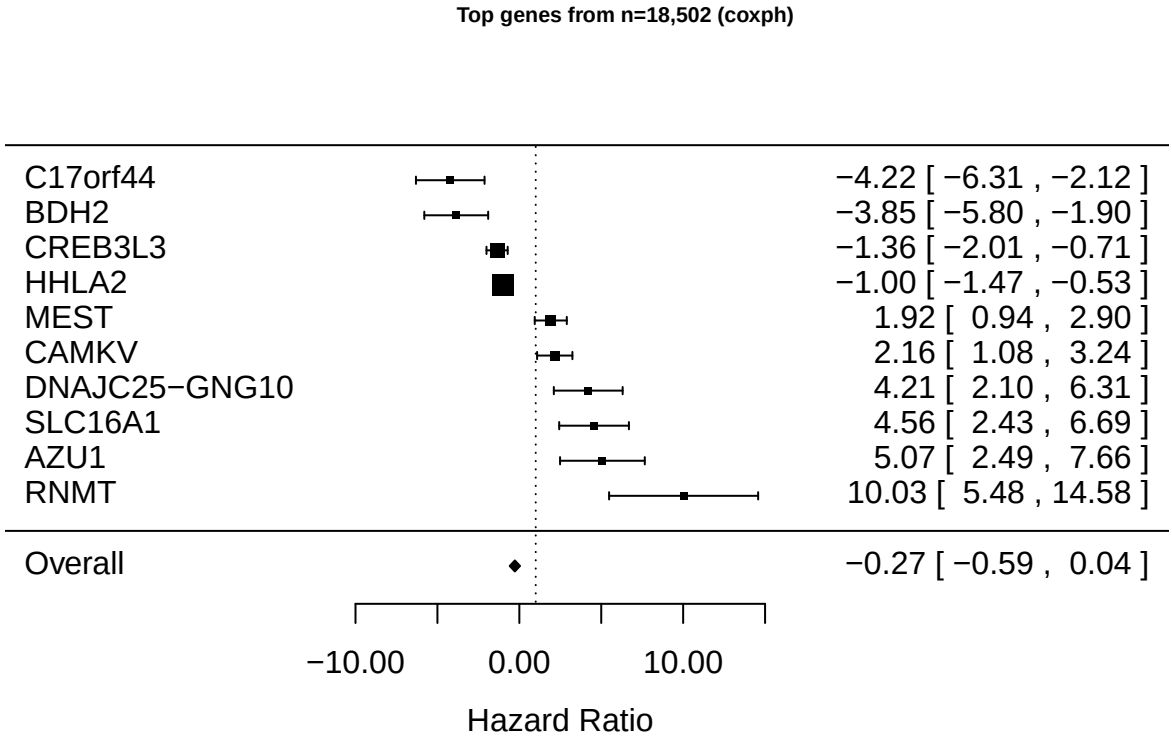
Survival Analysis of (all) genes in mRCC

I performed Coxph and Concordance index surv analysis to identify genes with prognostic value. I examined all genes (n=18,502).

If all genes are discretized into high/low expression, there were 0 genes that were significantly associated with outcome among all genes (n=18502), after correcting for multiple testing (BH, FDR). There were 1330 genes with unadjusted pvalues <0.05

If the continuous gene expression (coxph) model was used, there were 0 genes that were significantly (p adjust <0.05) associated with outcome among all genes (n=18502), after correcting for multiple testing (BH, FDR). There were 2107 genes with unadjusted pvalues <0.05





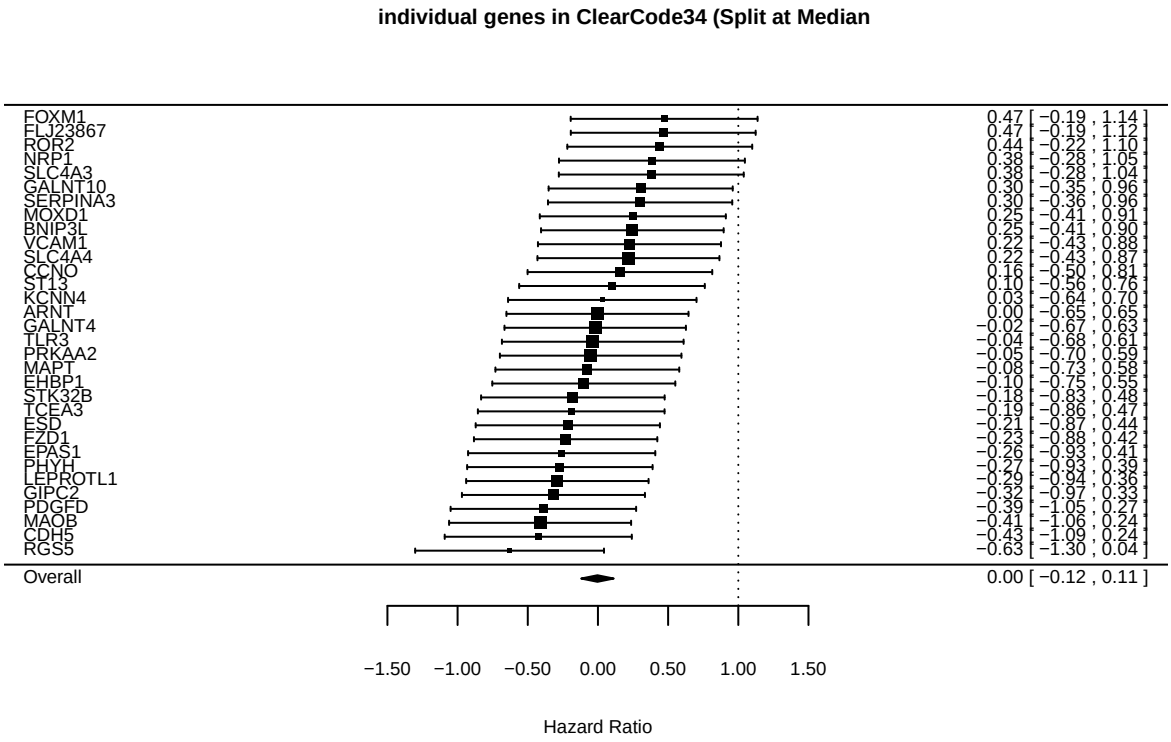
The gene ACACA was identified by the TCGA RCC paper. It was ranks 33 in the coxph analysis and 13 in the strafified by Median analysis

Prognostic power of ClearCode34 genes in mRCC

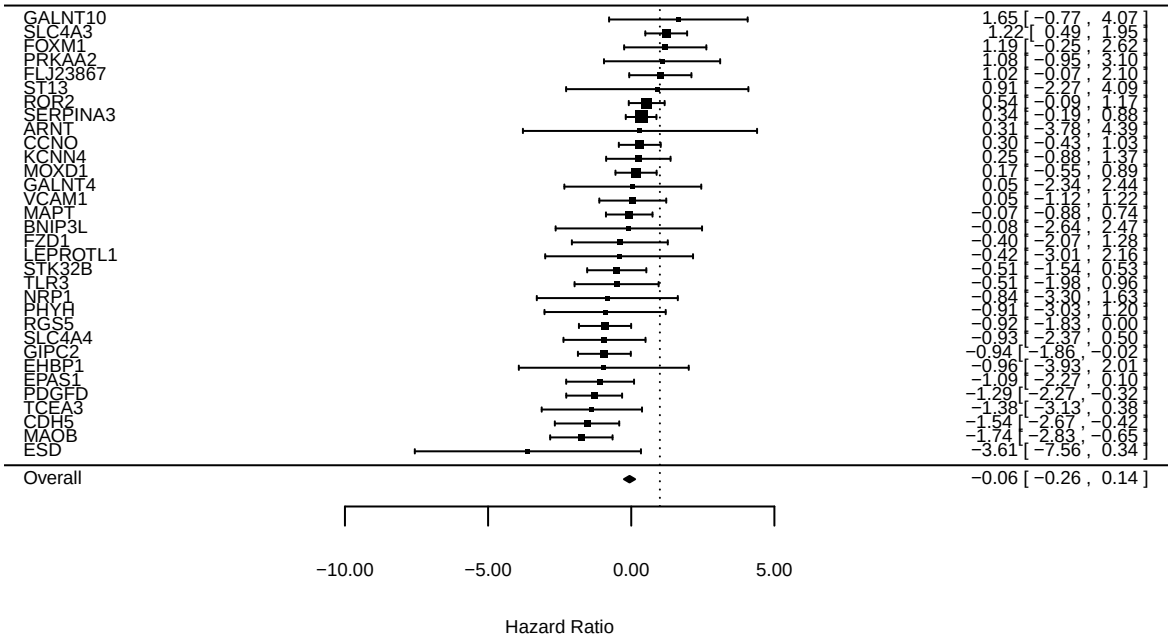
The prognostic performance of individuals ClearCode34 genes in mRCC (n=54) is given below.

Six genes were significant with unadjusted p-values <0.05 (SLC4A3, MAOB, CDH5, PDGFD GIPC2 and RGS5) in coxph analysis. The gene that was most significant was SLC4A3 (unadjusted pvalue p<0.01).

If I performed surv analysis on the statified gene expression by if it was below or above the median gene expression level. Then no genes were significant at p<0.05. The genes with lowest p-values were RGS5,FOXM1, FLJ23867 and ROR2.



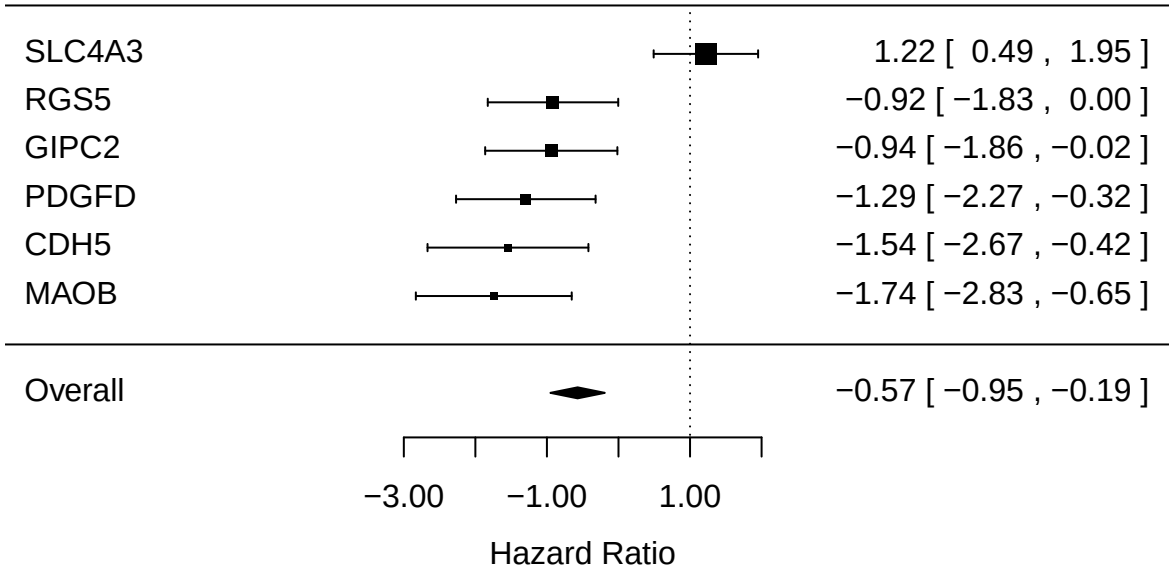
SYMBOL ENTREZID estimate std.error statistic p.value conf.low conf.high Score_logrank.p.adjust CI ——-



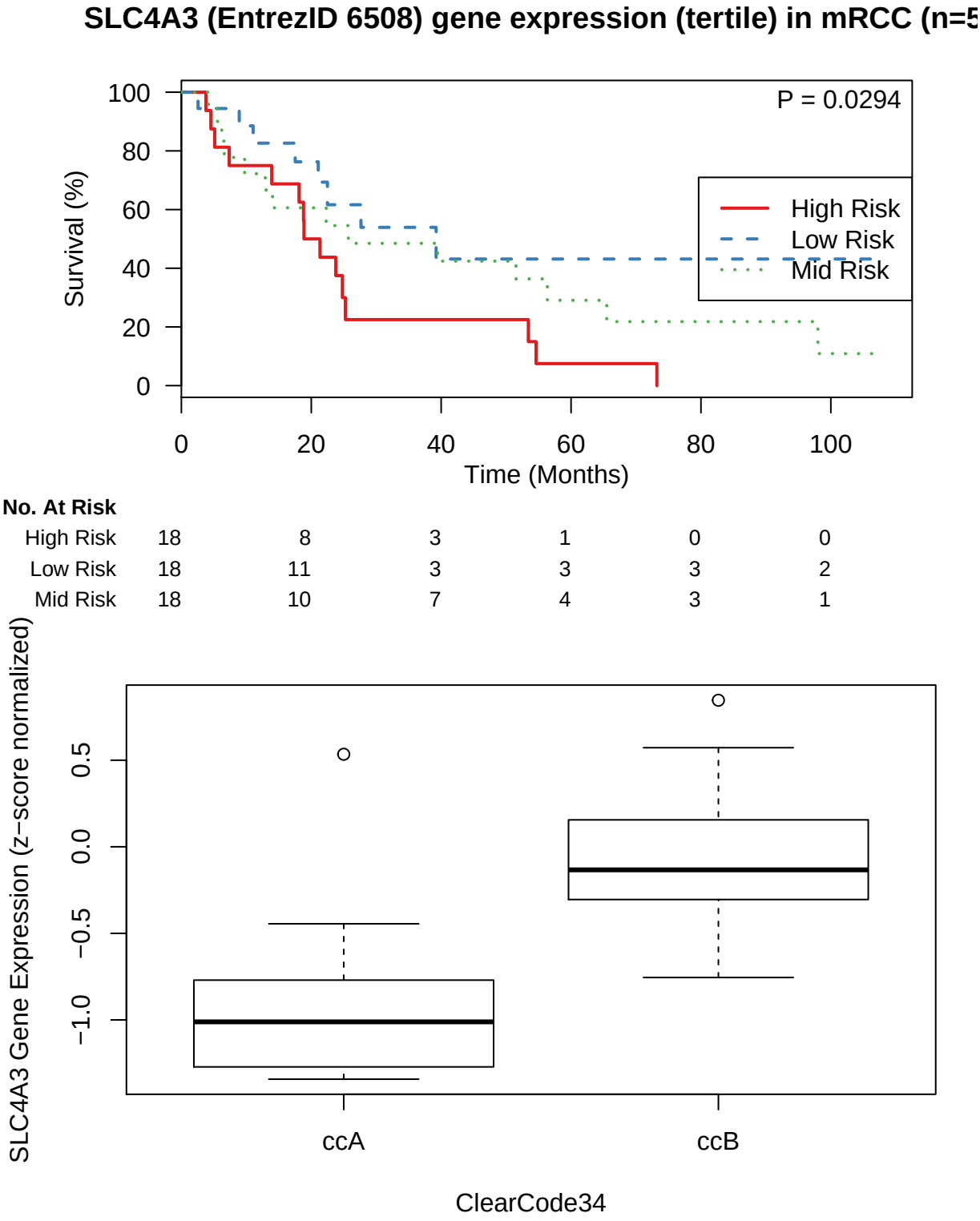
	SYMBOL	ENTREZID	estimate	std.error	statistic	p.value	conf.low	conf.high	Sco
SLC4A3|6508	SLC4A3	6508	1.2215281	0.3724171	3.280000	0.0010381	0.491604	1.9514523	
RGS5|8490	RGS5	8490	-0.9151381	0.4650388	-1.967875	0.0490824	-1.826597	-0.0036789	
GIPC2|54810	GIPC2	54810	-0.9391845	0.4711785	-1.993267	0.0462322	-1.862677	-0.0156917	

	SYMBOL	ENTREZID	estimate	std.error	statistic	p.value	conf.low	conf.high	Sco
PDGFD|80310	PDGFD	80310	-1.2949099	0.4974693	-2.602995	0.0092413	-2.269932	-0.3198880	
CDH5|1003	CDH5	1003	-1.5442733	0.5735738	-2.692370	0.0070946	-2.668457	-0.4200892	
MAOB|4129	MAOB	4129	-1.7426531	0.5556138	-3.136447	0.0017101	-2.831636	-0.6536701	

Genes in ClearCode34 (coxph) with p<0.05 (unadjusted)



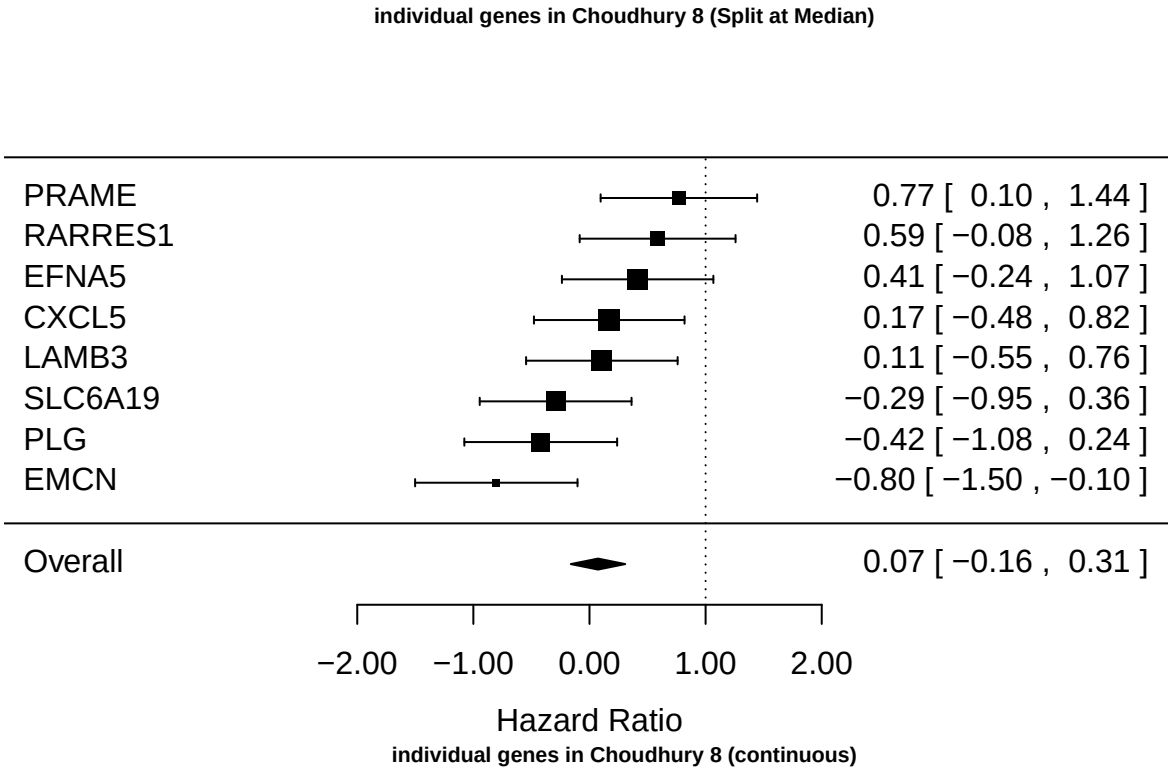

```
##  [1] SYMBOL                ENTREZID              estimate
##  [4] std.error             statistic             p.value
##  [7] conf.low              conf.high             Score_logrank.p.adjust
## [10] CI
## <0 rows> (or 0-length row.names)
```

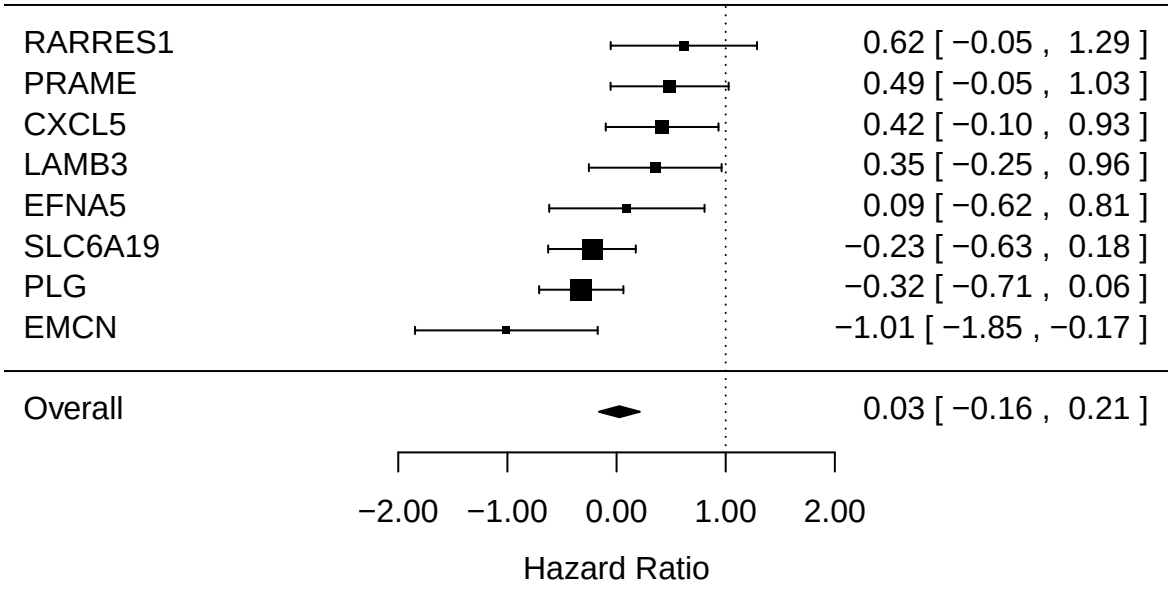


Prognostic power of Choudhury genes in mRCC

The prognostic ability of the Chodhury 8 genes in mRCC (n=54) is given below. I performed survival analysis on the entire (n=18502 genes), and none of the Choudhury genes genes were significant after correcting for multiple testing of 18,502 genes.

When I used the gene expression levels in the coxph, 1 gene had significant unadjusted p-values <0.05 (EMCN). Both it (EMCN) and PFRAME were also significant (unadjusted p-values <0.05) were gene expression values were cut at the median.





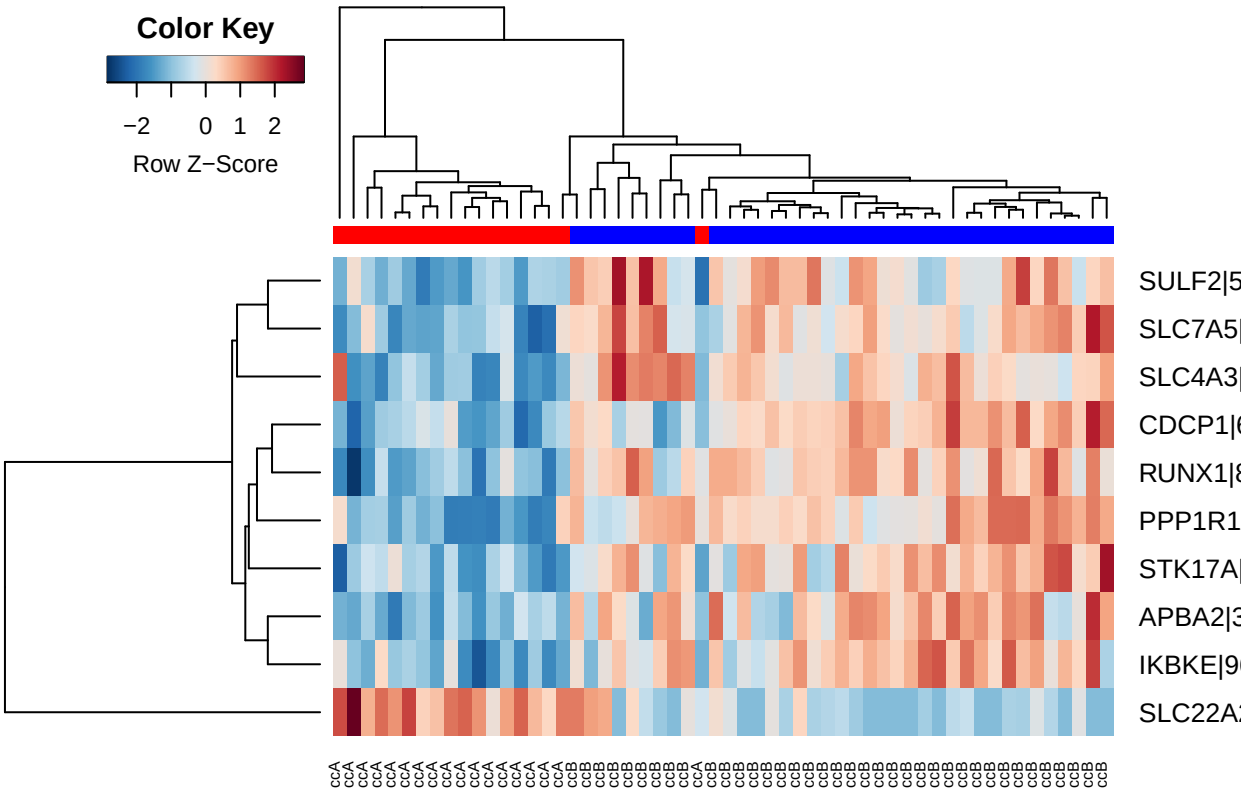
Expression of ClearCode and C8 genes in the ccA/ccB subtypes in mRCC

I examined whether the Choudhery C8 or ClearCode34 genes were differentially expressed between ccA/ccB in the 56 mRCC cases. I performed differential gene expression analysis to examined which of the 20531 genes where differentially expressed between the mRCC ccA and ccB cases.

The genes with greatest differential expression between ccA and ccB were RUNX1 and PPP1R1A.

	SYMBOL	ENTREZID	logFC	AveExpr	t	P.Value	adj.P.Val	B
RUNX1|861	RUNX1	861	0.4027089	1.0092633	9.669181	0	0e+00	20.14936
PPP1R1A|5502	PPP1R1A	5502	1.5481587	0.2631824	9.239994	0	0e+00	18.68269
SLC22A24|283238	SLC22A24	283238	-0.8462997	-1.0420243	-8.451478	0	1e-07	15.93698
SLC7A5|8140	SLC7A5	8140	0.6767131	0.9710850	8.237391	0	2e-07	15.18187
IKBKE|9641	IKBKE	9641	0.3653193	0.3103496	8.048144	0	3e-07	14.51158
SLC4A3|6508	SLC4A3	6508	0.8550086	-0.3431470	8.014083	0	3e-07	14.39069
SULF2|55959	SULF2	55959	0.3676699	1.3421734	7.979700	0	3e-07	14.26859
CDCP1|64866	CDCP1	64866	0.7672559	0.6343061	7.906579	0	3e-07	14.00869
APBA2|321	APBA2	321	0.5944995	-0.1515052	7.764376	0	5e-07	13.50240
STK17A|9263	STK17A	9263	0.2565321	0.7409249	7.683050	0	5e-07	13.21241

```
##      Class Color
## [1,] "ccA" "red"
## [2,] "ccB" "blue"
```

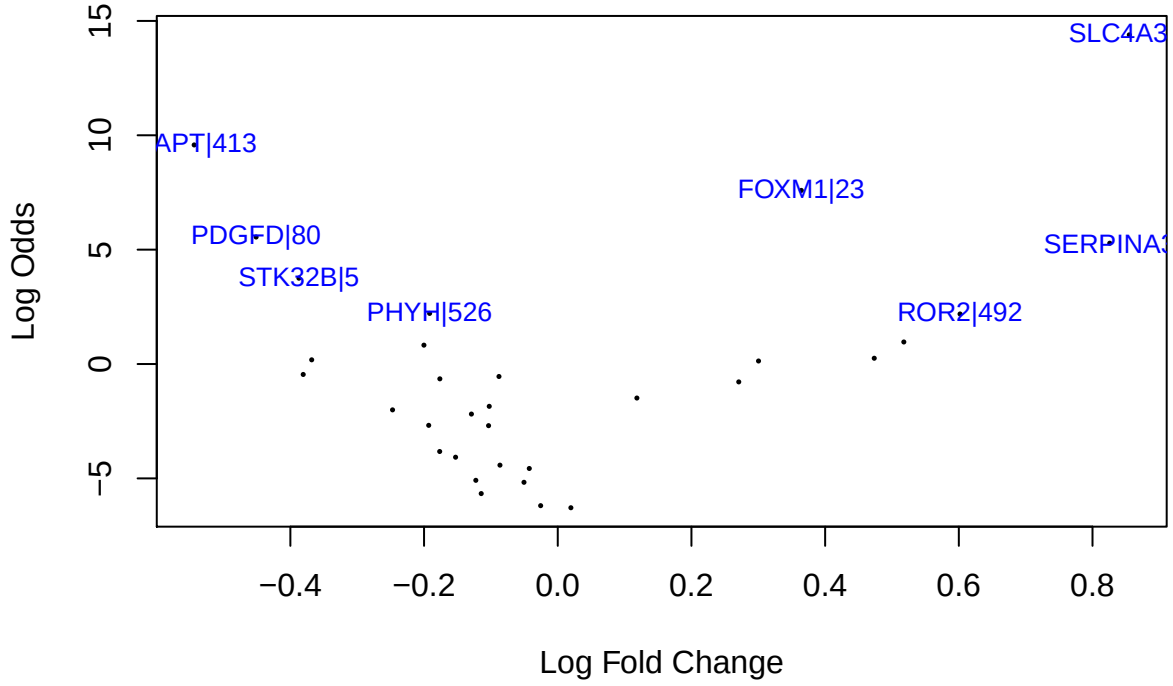


Most of 23/ 34 genes in the ccA/ccB signature and many of the C8 signature (5/8) were differentially expressed (p.adjust <0.05) between ccA/ccB.

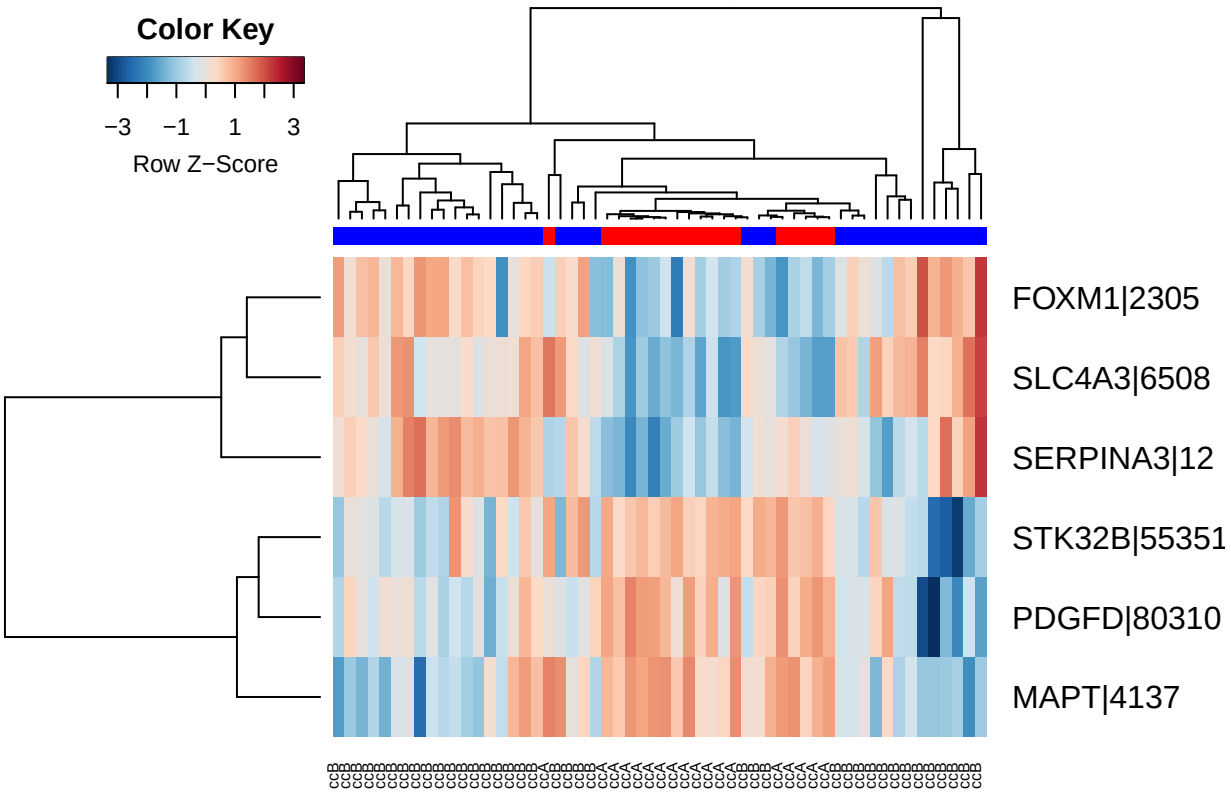
Among the Clearcode genes, upregulation of SLC4A3, SERPINA3 and FOXM1 and down regulation of MAPT, STK32B and PDGFD in ccB were among the genes with highest differential expression. The heatmap

below only shows genes with significant p value less than 0.001 (adjusted p values). SLC4A3 was ranked the 8th most differentially expressed gene between ccA/ccB. MAPT, FOXM1, PDGFD, SERPINA3 were ranked 38th, 80th, 161th and 179th respectively.

```
##
## FALSE  TRUE
##     9    23
```

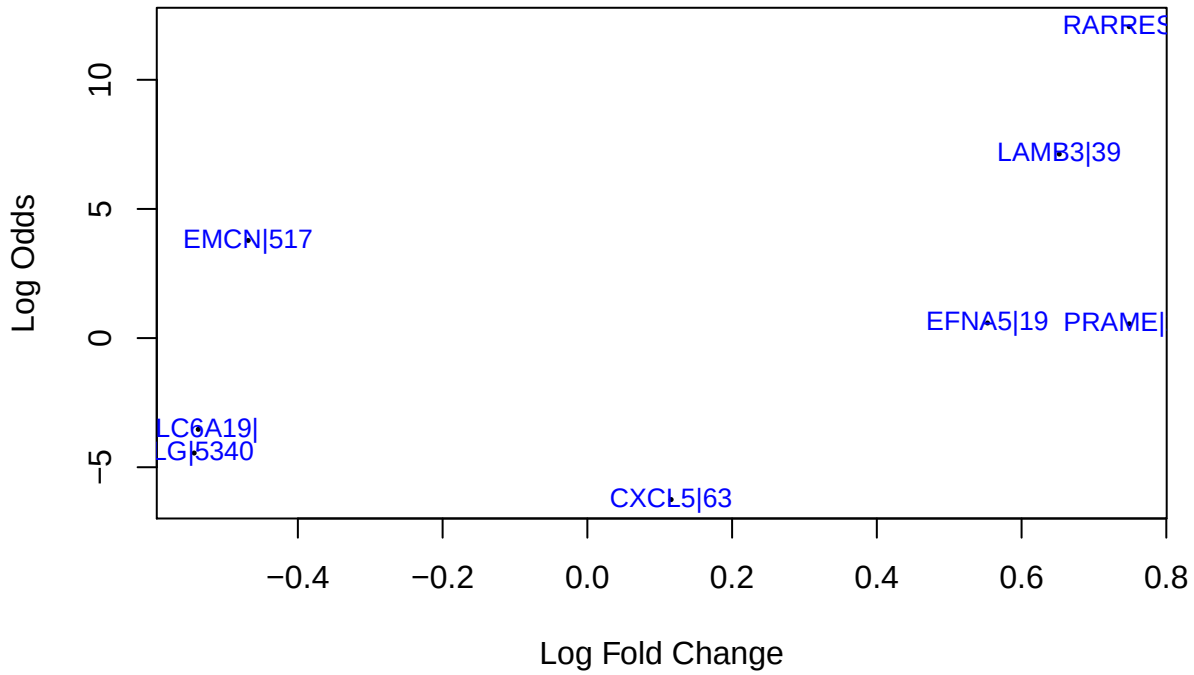



```
##      Class Color
## [1,] "ccA" "red"
## [2,] "ccB" "blue"
```

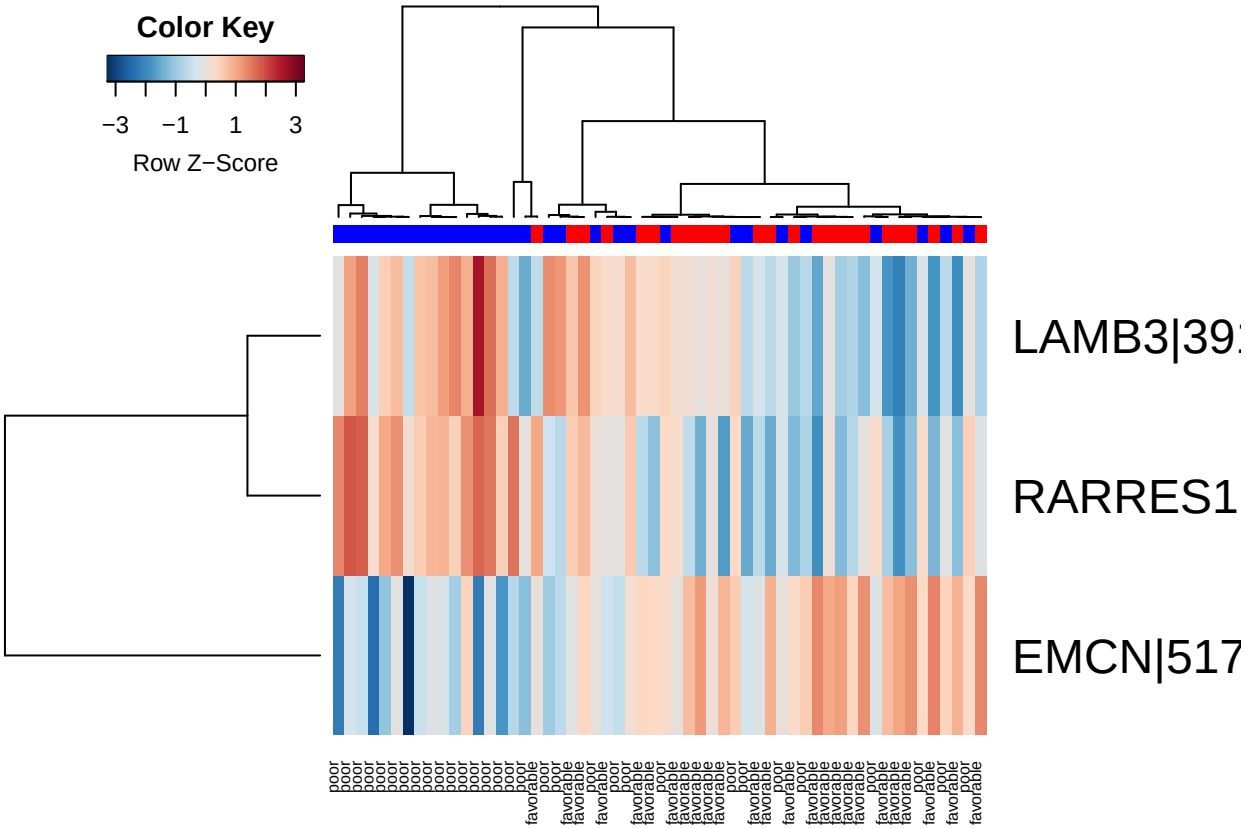


Among the C8 genes, upregulation of LAMB3 and RARRES1 and down regulation of EMCN in ccB were among the genes with highest differential expression. The heatmap below only shows the 3 genes with significant p value less than 0.001 (adjusted p values). RARRES1 wanted ranked 18th most differentially expressed between ccA/ccB. LAMB2 was 97th, and EMCN was 306th. The other genes had a rank of 900 to 15,353 among the 20531 genes examined

```
##
## FALSE  TRUE
##     3     5
```




```
##       Class       Color
## [1,] "favorable" "red"
## [2,] "poor"      "blue"
```



More comparision between signatures- Significant Pathways or Go Terms (** UN ** adjusted p values)

Since there were few significant pathways that were significant in the C8 gene signatured I explored genesets and pathways that had insignificant adjusted p.values

Number of genes with significant unadjusted p.values

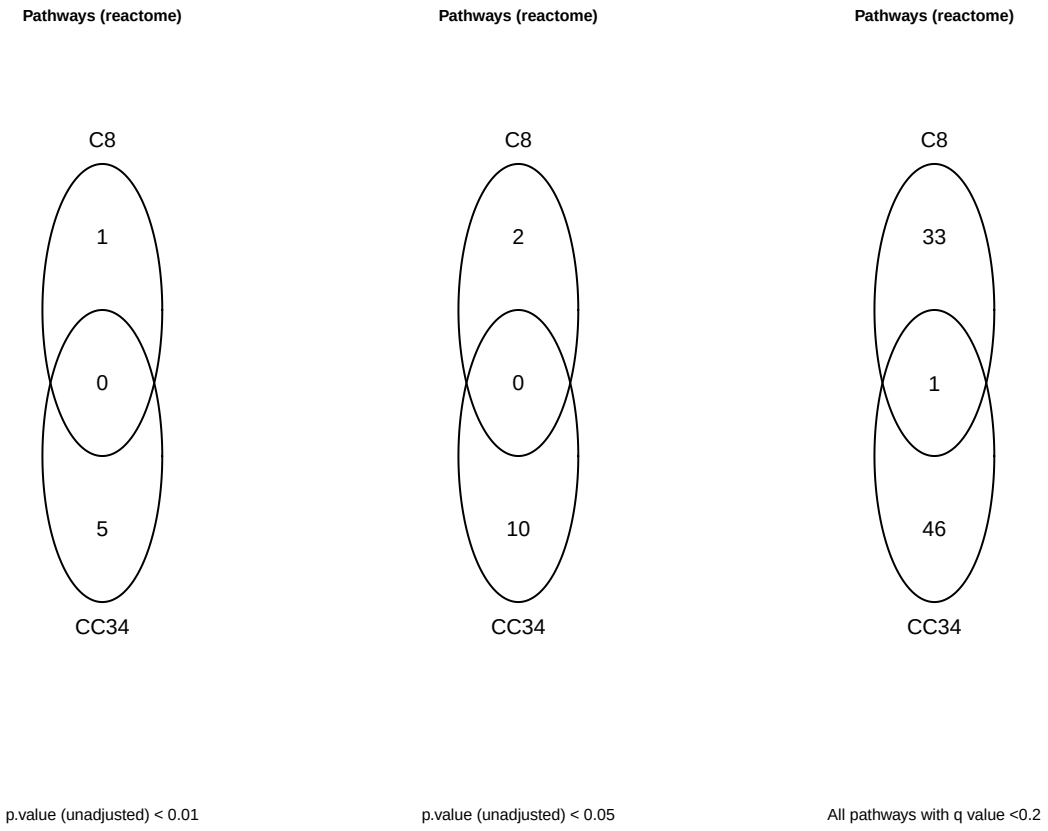
	Pval0.05	Pval0.01	Pval0.001	Pval1e-04
CC34.Pathway	10	5	2	0
C8.Pathway	2	1	0	0
CC34.GO_BP	180	72	12	0
C8.GO_BP	10	7	1	0
CC34.GO_MF	20	15	5	1
C8.GO_MF	1	1	0	0
CC34.GO_CC	0	0	0	0
C8.GO_CC	2	1	0	0

Comparision of Pathway Enrichment

I compared the overlap in highly ranked Reactome pathways. There were 10 and 2 reactome pathways there were enriched in genes in the Clearcode signature and Choudhury signature respectively (unadjusted $p<0.05$,).

When I relaxed the criteria and considered unadjusted p.values, in which there was 2 gene overlap. , there were genes from Clearcode and Choudhury were enriched in 5 and 1 reactome pathways ($p<0.01$, unadjusted) and 10 and 2 pathways ($p<0.05$, unadjusted)

There was no overlap among these in pathways with unadjusted $p<0.05$ or 0.01. If I ignore pvalue, and expand the ranks to include all pathways with a 20% FDR only 1 pathway overlap. Reactome Pathway "425393", Transport of inorganic cations/anions and amino acids/oligopeptides was ranked 14th by CC34 and 23rd by C8 and was due to genes SLC4A4/SLC4A3 in CC34 and SLC6A19 in C8.



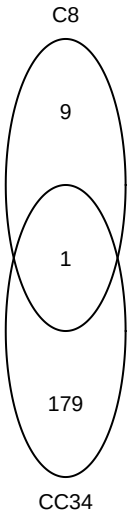
Comparision of Gene Ontology Enrichment Results: Biological Process

Neither Clearcode34 or Choudury genes were significantly enriched in any biological process GO terms when the p values were adjusted for multiple testing (FDR).

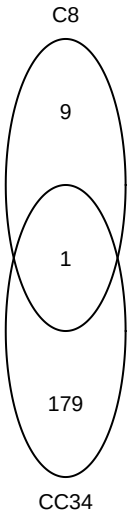
In **uncorrected** p.values <0.05, there were 180 and 10 Gene Ontology Biological processes that were enriched in Clearcode or Choudhury genes respectively.

I compared these and the top 100 Gene Ontology (biological prcesss) terms from each signature. There was little overlap in Biological Process.

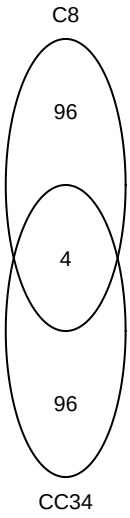
GO: Biological Process (P<0.05, Count>1)



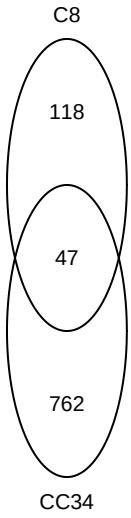
GO: Biological Process (P<0.05, Count>1)



GO: Biological Process (top 100)



GO: Biological Process (qvalue<0.2)

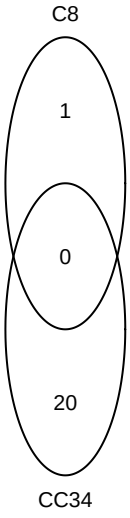


Comparision of Gene Ontology Enrichment Results:Molecular Function

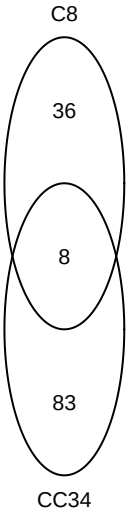
In **uncorrected** p.values <0.05, there were 20 and 1 Gene Ontology Biological processes that were enriched in Clearcode or Choudhury genes respectively.

I compared these and the top 100 Gene Ontology (biological prcesss) terms from each signature. There was little overlap in Molecular Function terms.

GO: Molecular Function (P<0.05, Count>1)



GO: Biological Process (qvalue<0.2)



Only 1 mTOR pathway genes found in clearCode34 (none in Choudury signature)

Within reactome there are three pathways with the word "mTOR" in the title, I looked at these genes, but found only 1 overlapped with ClearCode34. There was no overlap Choudhury 8 gene signature. The gene PRKAA2 was in the ClearCode34 signature and in the mTOR pathways

```
## [1] "Energy dependent regulation of mTOR by LKB1-AMPK"
## [2] "mTOR signalling"
## [3] "mTORC1-mediated signalling"
```

```
## $`Energy dependent regulation of mTOR by LKB1-AMPK`
##  [1] "LAMTOR5" "PRKAB2"  "PPM1A"   "MTOR"    "TSC2"    "PRKAG1"  "PRKAA2"
##  [8] "PRKAA1"  "RHEB"    "STK11"   "RRAGB"   "LAMTOR1" "RRAGA"   "STRADA"
## [15] "RPTOR"   "TSC1"    "MLST8"   "STRADB"  "CAB39L"  "RRAGC"   "RRAGD"
## [22] "PRKAG3"  "PRKAG2"  "LAMTOR3" "LAMTOR2" "CAB39"   "PRKAB1"  "LAMTOR4"
##
## $`mTOR signalling`
##  [1] "LAMTOR5"      "PRKAB2"       "EIF4E"        "RPS6KB1"
##  [5] "AKT1"         "AKT2"         "YWHAB"        "PPM1A"
##  [9] "MTOR"         "TSC2"         "PRKAG1"       "PRKAA2"
## [13] "LAMTOR4"      "PRKAA1"       "RHEB"         "STK11"
## [17] "RRAGB"        "LAMTOR1"      "RRAGA"        "STRADA"
## [21] "RPTOR"        "TSC1"         "AKT1S1"       "MLST8"
## [25] "STRADB"       "CAB39L"       "RRAGC"        "RRAGD"
## [29] "PRKAG3"       "PRKAG2"       "LAMTOR3"      "LAMTOR2"
## [33] "CAB39"        "PRKAB1"       "EIF4EBP1"     "EEF2K"
## [37] "LOC101930123" "EIF4B"        "RPS6"         "EIF4G1"
##
## $`mTORC1-mediated signalling`
##  [1] "LAMTOR5"      "EIF4E"        "RPS6KB1"      "YWHAB"
##  [5] "MTOR"         "LAMTOR4"      "RHEB"         "RRAGB"
##  [9] "LAMTOR1"      "RRAGA"        "RPTOR"        "AKT1S1"
## [13] "MLST8"        "RRAGC"        "RRAGD"        "LAMTOR3"
## [17] "LAMTOR2"      "EIF4EBP1"     "EEF2K"        "LOC101930123"
## [21] "EIF4B"        "RPS6"         "EIF4G1"
```

```
##      Energy dependent regulation of mTOR by LKB1-AMPK mTOR signalling
## C8                                                  0               0
## CC34                                                1               1
##      mTORC1-mediated signalling
## C8                            0
## CC34                          0
```

The genes PRKAA2 (Entrez gene ID 5563) is involved in mTor Signalling and was in the ccA/ccB signature. There was none of these mTor genes in the Choudhury et al., 8 gene signature

```
## $`Energy dependent regulation of mTOR by LKB1-AMPK`
## PRKAA2
## "5563"
##
## $`mTOR signalling`
## PRKAA2
## "5563"
##
## $`mTORC1-mediated signalling`
```

```
## named character(0)
```